

 FACTS ON FILE MATH LIBRARY

THE HISTORY OF
MATHEMATICS

Numbers



COMPUTERS,
PHILOSOPHERS & THE
SEARCH FOR MEANING

JOHN TABAK, PH.D.

NUMBERS

COMPUTERS, PHILOSOPHERS,
AND THE SEARCH FOR MEANING



THE HISTORY OF MATHEMATICS

NUMBERS

COMPUTERS, PHILOSOPHERS,
AND THE SEARCH FOR MEANING

John Tabak, Ph.D.



Facts On File, Inc.

NUMBERS: Computers, Philosophers, and the Search for Meaning

Copyright © 2004 by John Tabak, Ph.D.
Permissions appear after relevant quoted material.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval systems, without permission in writing from the publisher. For information contact:

Facts On File, Inc.
132 West 31st Street
New York NY 10001

Library of Congress Cataloging-in-Publication Data

Tabak, John.

Numbers: computers, philosophers, and the search for meaning / John Tabak.

p. cm. — (History of mathematics)

Includes bibliographical references and index.

ISBN 0-8160-4955-6 (acid-free paper)

1. Numeration—History 2. Counting—History. I. Title.

QA141.T33 2004

513.2—dc222003016970

Facts On File books are available at special discounts when purchased in bulk quantities for businesses, associations, institutions, or sales promotions. Please call our Special Sales Department in New York at (212) 967-8800 or (800) 322-8755.

You can find Facts On File on the World Wide Web at <http://www.factsonfile.com>

Text design by David Strelecky
Cover design by Kelly Parr
Illustrations by Patricia Meschino

Printed in the United States of America

MP FOF 10 9 8 7 6 5 4 3 2 1

This book is printed on acid-free paper.

To Sandra and James, world travelers, good company.

CONTENTS

Acknowledgments	x
Introduction: Number and Imagination	xi
Part One: Numbers for Computation	I
1 The First Problems	3
2 Early Counting Systems	11
<i>A Mesopotamian Education</i>	13
The Mesopotamian Number System	14
<i>Mesopotamian Mathematics Homework</i>	18
The Egyptian Number System	20
<i>A Problem from the Ahmes Papyrus</i>	24
The Mayan Number System	26
The Chinese Number System	30
<i>A Problem from the Nine Chapters</i>	32
3 Our Place Value Number System	34
<i>Explaining the New System</i>	38
4 Analytical Engines	46
<i>Calculators, Computers, and the Human Imagination</i>	48
Charles Babbage and the Analytical Engine	50
<i>An Early Electronic Representation of Our Number System</i>	52
Floating-Point Representation	56
<i>Floating-Point Arithmetic and Your Calculator</i>	58
<i>Why Computers?</i>	60

Part Two: Extending the Idea of a Number 65

5 An Evolving Concept of a Number	67
Irrational Numbers	70
Pythagoras of Samos	73
<i>The Irrationality of $\sqrt{2}$</i>	75
6 Negative Numbers	77
<i>Ancient Mathematical Texts from the Indian Subcontinent</i>	82
<i>Out of India</i>	83
7 Algebraic Numbers	85
Tartaglia, Ferrari, and Cardano	88
Girard and Wallis	93
Euler and d'Alembert	97
<i>The Debate over “Fictitious” Numbers</i>	98
The Complex Numbers: A Modern View	103
<i>Using Complex Numbers</i>	104
8 Transcendental Numbers and the Search for Meaning	107
Dedekind and the Real Number Line	111

Part Three: The Problem of Infinity 117

9 Early Insights	119
10 Galileo and Bolzano	130
<i>Infinity as a Number</i>	136
The Life and Opinions of Tristram Shandy, Gentleman	140
11 Georg Cantor and the Logic of the Infinite	143
<i>There Are No More Rational Numbers than Natural Numbers</i>	146
<i>There Are More Real Numbers than Natural Numbers</i>	148

The Russell Paradox	154
Resolving the Russell Paradox	159
12 Cantor's Legacy	165
Kurt Gödel	170
<i>Formal Languages Today</i>	173
Alan Turing	174
Chronology	181
Glossary	199
Further Reading	203
Index	213

ACKNOWLEDGMENTS

The author is deeply appreciative of Frank K. Darmstadt, Executive Editor, for his many helpful suggestions and of Dorothy Cummings, Project Editor, for the insightful work that she did editing this volume.

Special thanks to Penelope Pillsbury and the staff of the Brownell Library, Essex Junction, Vermont, for their extraordinary help with the many difficult research questions that arose during the preparation of this book.

INTRODUCTION

NUMBER AND IMAGINATION

It is not easy to appreciate just how difficult and how interesting counting can be. We do not think about counting; we just count. Faced with a collection of objects we often cannot help counting them. Cars, shoes, coins, or people—we count them all. Counting is, after all, one of the first things we learn. Counting is a basic life skill.

Counting is deceptive. It is so easy to count that a child can learn to do it, but ease in counting is a fairly recent development. Throughout most of human history counting was done by specialists, because counting was hard work. It took thousands of years to develop the counting system that we use today. Throughout history brilliant people in different places and different cultures have contributed to our current method of counting. Vital contributions were made in Mesopotamia and India and other places besides. The method we now use to count and write numbers is so simple and so versatile that most of us never feel the need to learn any other method. Our way of counting is powerful enough to handle almost any situation.

There has never been just one way of counting. Before humankind was so tightly linked by ship and plane, phone and computer, different cultures developed different ways of counting. Some of these methods were abandoned long ago, but others, especially the ones used in computers, have only been widely adopted in the last 50 or 60 years, although the methods themselves were known centuries before the invention of the transistor. The history of counting is as diverse as it is ancient.

Numbers are, however, about more than counting. Since the days of the ancient Greeks mathematicians have identified several different types of numbers. Each type of number has properties that are unique to it. More often than not, the discovery of a new

type of number has generated controversy, and acceptance of a new kind of number has never been quick. This was true in ancient Greece, and it was true 2,000 years later during the Enlightenment.

The difficulty that people have in accepting new kinds of numbers has no relation to the numbers themselves, of course. In a sense the numbers were there all the time, waiting to be discovered. The difficulty lies in our imagination. It takes time to overcome the restrictions *that we place on our own ideas of what numbers are*. Initially new kinds of numbers were often described by what they are not. Irrational numbers were initially described as numbers that are not rational. (A rational number can be written as a fraction with whole numbers in the numerator and denominator.) Imaginary numbers were at first described as numbers that are not real. (A real number has a place on the real number line.) Developing the imagination and insight necessary to describe these “new” numbers in terms of what they are, instead of what they are not, took centuries. This was finally accomplished less than 150 years ago. We will see how this was done.

Finally, we will investigate infinity. Infinity is an idea that philosophers and mathematicians, curious laypeople, and even curious children have pondered for millennia. In mathematics questions about the infinite have often been expressed in terms of sets of numbers. For example how many counting numbers are there? The easy answer, of course, is that there are infinitely many counting numbers. But that is not a very satisfactory answer, because it tells us nothing about what “infinitely many” means. For millennia mathematicians knew only what infinite sets are not: Infinite sets are not finite. This definition offers no insight into what infinite sets are. Specifically what are infinite sets? Are some infinite sets larger than others? (For example are there more fractions than there are counting numbers?) Can we define “infinitely large” numbers to represent the size of infinite sets in the same way that the number 1, for example, represents the size of any set with just a single element? If we can define infinitely large numbers, how many such “infinite numbers” exist? Is there an arithmetic of the infinite? These questions were asked centuries ago, but there

were few satisfying answers until the latter half of the 19th century, because until relatively recently no one truly knew what an infinite set is. Progress, however, has been made. Mathematicians now understand the infinite in a positive way. There are rigorous answers to the questions posed earlier in this paragraph, and the answers are often surprising and surprisingly useful. In this book we look at the profound impact that such abstract thinking has had inside and outside mathematics.

We begin, however, at the beginning.

PART ONE

NUMBERS FOR COMPUTATION

1

THE FIRST PROBLEMS

It would be interesting to know when people first began to develop a concept of number. There are no cases of a culture that developed a method of writing words without also—and simultaneously—developing a method of writing numbers. Just as they knew how to speak before they began to write they must also have developed a concept of number before they developed a written language. As a consequence written records are of limited use in tracing the earliest ideas about number. The records do not go far enough back in time, and the prehistoric evidence is inconclusive. Scientists have unearthed stones from prehistoric sites that were carefully painted in decidedly nonrandom patterns. They have also found animal bones with numerous notches, carefully grouped and carefully carved, cut into the bones. Are these early attempts at recording numbers or are they just decorations? Without a “user’s manual” it is difficult to know for sure.

It seems likely that early humans could compare quantities long before they could count them. There is often less technique involved in comparing two sets than in counting the elements of either. To understand how this applies to numbers, consider a common design for an early odometer, a device for measuring distances. Leonardo da Vinci designed one such device. It looked a lot like a wheelbarrow: The device had a single front wheel and the user moved it from behind by lifting the handles and pushing. The front wheel was attached through a series of gears to a circular device that revolved at the top of the odometer. This circular device was filled with pebbles. On the side of the odometer beneath the container of pebbles was a box. As the odometer was



These painted pebbles were excavated from the cave of Mas d'Azil, Ariège, France. Approximately 1,400 painted pebbles have been discovered at this site. They were painted about 11,000 years ago. Their meaning is unknown. (© Copyright The British Museum)

pushed forward the wheel rotated, and that rotation in turn caused the circular container of pebbles to turn. As the container of pebbles rotated it dropped a pebble—one pebble at a time—into the side box. Because the whole machine was geared to the wheel, each unit of distance traveled was represented by one pebble in the box. The farther the machine was pushed, the more pebbles dropped into the box. This measurement device established a *one-to-one correspondence* between the distance traveled and the number of pebbles in the box. There were exactly as many pebbles in the box as there were units of distance traveled by the odometer. Once the user knew the number of pebbles in the box, she or he also knew the distance traveled by the odometer. It is a simple idea, and it has far reaching implications for higher mathematics, but in itself Leonardo's concept of measurement tells the user little about the distance traveled by the odometer.

Although it may seem that Leonardo's odometer solves the problem of how to measure distances along a road, it does not.

In fact it just shifts the problem from one of counting feet (or meters) to one of counting pebbles. Leonardo's device establishes a correspondence between the number of units of distance traveled and the number of pebbles in the box, but at the end of the day, all the user really knows is that there are as many pebbles in the box as there are units of distance traveled. If you have ever tried to guess how many jelly beans are in a large full jar just by looking then you know that estimates can vary widely from one person to the next. It is rare for anyone to guess the number of beans in the jar correctly without emptying out the contents and counting the beans by hand.

Despite its shortcomings the simple idea of establishing a correspondence between one group of objects and another is valuable because it enables us to establish a clear definition of what a number is. In fact the idea that there must exist a one-to-one correspondence between any two sets of the same size—and that this correspondence is precisely the way we know that the two sets are the same size—lies at the heart of modern ideas about number. At some level it lies at the heart of all ideas about number, and so this abstract idea must have been formulated early in human history. People had to discover that five stones and five fingers have something in common. They had to discover that five stones and five fingers are both *instances* of the number 5. Notice that



Leonardo da Vinci, self-portrait, ca. 1500. His odometer established a one-to-one correspondence between units of distance traveled and stones deposited in a box. (Science Museum, London/Topham-HIP/The Image Works)

6 NUMBERS

there is a clear distinction between examples or instances of the number 5 and the number 5 itself. There is only one number 5, but we are surrounded by many examples of the number 5. The *concept* of a number is a much deeper idea than an *example* of a number.

In the 20th century the eminent mathematician Bertrand Russell wrote an entire article, “The Definition of Number,” in the hope of providing an unambiguous description of the concept of number. The fact that it took a highly creative mathematician to write such an article indicates that the subject is not as straightforward as it may at first seem. Russell wrote that the number 5, for example, *is* a set consisting of a collection of sets. Each set in the collection consists of five elements, so each set is an instance of the number 5. There is no set in the collection that is not an instance of the number 5, and all instances of the number 5 are included in the collection. In particular the set consisting of the letters *a*, *b*, *c*, *d*, and *e* belongs to the number 5. So does the set consisting of the states Alaska, Maine, South Carolina, New York, and Wisconsin. Many other five-element sets are included as well, of course. Describing a number as the collection of all instances of that number may seem awkward, even strange, but it has proved to be a very useful idea. Russell’s definition, which we examine more closely later, arose as mathematicians worked to understand and classify sets that are infinite as well as finite. On some level his idea must have been shared by even the earliest of mathematicians who needed to recognize the equivalence of all sets with the same number of elements.

Having established a concept of number, the earliest mathematicians needed to find a way to record numbers. Part of what makes numbers so useful is that they enable us to classify sets of objects according to size. Ten bushels of grain is better than five. An army of 10,000 soldiers has the advantage over an army of 100 soldiers all other factors being equal. To *use* numbers to describe the world around them these ancient peoples had to develop a method of counting. This was the second mathematical barrier to progress.

A nice historical example of a fairly modern counting method was used in Madagascar for the purpose of counting soldiers. Here

is how it worked: The soldiers walked single file through a narrow passage or gate. As each soldier passed through the gate, a pebble was dropped into the 1s pile. Each pebble in this pile represented one soldier. With every 10th soldier the 1s pile was cleared away and an additional pebble was dropped in the 10s pile. The pattern was then resumed. Each pebble in the 10s pile represented 10 soldiers. When there were 10 pebbles in the 10s pile, that pile, too, was wiped clean and a single additional pebble added to the 100s pile. Notice that unlike Leonardo's odometer, in which every pebble represents one unit of distance, in the Malagasy scheme no pebble has a fixed value. The value a pebble acquires depends on the pile in which it is placed.

One advantage of the Malagasy method of counting is that "pebblewise" it is very economical. Whereas Leonardo's odometer, for example, would have required 123 pebbles to represent the number 123, the counters of Madagascar would have required only six pebbles: one pebble in the 100s pile, two in the 10s pile, and three in the 1s pile. In fact using the Madagascar idea, any counting number less than 100 can be represented by no more than 18 pebbles. (We never need more than nine pebbles in the 1s pile and nine pebbles in the 10s pile.) Similarly we need no more than three piles of nine pebbles to represent any counting number less than 1,000—nine in the 1s, nine in the 10s, and nine in the 100s—for a total of 27 pebbles. The larger the number, the more impressive the Malagasy approach appears. For example, any counting number less than 1 million can be represented by no more than 54 pebbles. This method of counting made it easy to keep track of large quantities of everything from soldiers to grain. It was an important conceptual breakthrough.

Notice that in the Malagasy scheme for counting soldiers the number 10 plays a special role. Each pile—the 1s pile, the 10s, the 100s, and so forth—can be in only one of 10 "states," where the word *states* refers to the number of stones in the pile. The pile may be empty—the "zero state"—or it may contain up to nine stones. There are no other possibilities. We describe this situation by saying that the Malagasy scheme is a base 10 scheme. It is one of the simplest and purest representations of counting in base 10 that we

8 NUMBERS

can imagine, but it can easily be modified so that one can count in any other base. In base 5, for example, the number 5 plays a special role: As the first soldier walked by, the counter would place a single stone in the 1s pile. Again each pebble in this pile represents one soldier. When there are five stones in the pile, the 1s pile is wiped clean and a single stone placed in the 5s pile. Each stone in the 5s pile represents five soldiers. When there are five stones in the 5s pile that pile is wiped clean and a single stone is placed in the 25s pile ($25 = 5 \times 5$). Each stone in this pile represents 25 soldiers. The stones are the same, of course, but the *meaning* of the stone again depends on the pile in which it is placed. In base 5 the pattern continues into the 125s pile ($125 = 5 \times 5 \times 5$), and when the 125s pile has five stones in it, it, too, is wiped clean and a single stone is placed in the 625s pile ($625 = 5 \times 5 \times 5 \times 5$). If we wanted to count in base 20 the same sort of technique would be used. The difference is that in this case we would have a 1s pile, a 20s pile, a 400s pile—because $400 = 20 \times 20$ —and beyond the 400s pile there would be an 8,000s pile—because $8,000 = 20 \times 20 \times 20$ —and so on.

Conceptually we do exactly the same thing when we count in our number system. We do not use stones; we use numerals, but the principle is the same. Because we count in base 10, our “piles” represent powers of 10: ones, tens, hundreds, and so on. We seldom, if ever, count in a base other than 10, but there is no particular advantage to counting in base 10, and in some circumstances it is distinctly disadvantageous to do so. There are many instances in which mathematicians have found it convenient for one reason or another to count in a base other than 10. The Mayans of Central America, for example, counted in base 20. The Mesopotamians, as we soon see, counted in base 60. The chart to follow is modeled on the stone piling technique just described to show how different numbers would look in the bases 5, 10, and 20.

It is interesting to note that some cultures never developed a systematic method of counting. These same cultures often developed complex mythologies and varied forms of artistic expression, but they required few if any numbers to express their

Num- ber	Base 10			Base 5				Base 20		
	100's	10's	1's	125's	25's	5's	1's	400's	20's	1's
7			◆◆◆◆			◆	◆			◆◆◆◆
26		◆	◆◆◆		◆		◆		◆	◆◆◆
123	◆	◆	◆◆		◆◆	◆◆	◆◆		◆◆◆◆	◆◆
500	◆◆◆			◆◆				◆	◆◆◆	

This chart shows how the numbers 7, 26, 123, and 500 can be represented in bases 10, 5, and 20.

thoughts. Some groups of aboriginal Australians are said to have lived almost without numbers until recent historical times. Early anthropologists indicated that some groups of aboriginal Australians did not distinguish numbers much beyond 6 or 7, and that counting proceeded along the lines of “1, 2, 3, 4, 5, 6, many.” We have to be careful in interpreting these early reports. Cultural and linguistic differences made it difficult for the anthropologists to communicate effectively with the people whom they studied, and their observations cannot be repeated since aboriginal Australians live in a different culture now and have adapted accordingly. In any case there is little doubt that not every culture developed a systematic method of counting. Not every culture developed a method of counting that enabled the user to count into the thousands or the millions or even higher.

At different times in history sometimes independently of each other, people in various widely separated cultures began to come to terms with the idea of number. They developed a concept of number, and they made the first steps toward expressing numbers in a base that suited their needs. This slow evolution

10 NUMBERS

of the most rudimentary concepts about numbers occupied the greater part of the prehistory of the human race. The situation began to change, however, as people lived together in cities and together strove for the common good. There was an increased need for record keeping and for a system that enabled the user to perform ever more sophisticated computations. Progress began to accelerate.

2

EARLY COUNTING SYSTEMS

Mesopotamian civilization was centered in the area of the Tigris and Euphrates Rivers in what is now the country of Iraq. The inhabitants of this region invented one of the oldest of all written languages, and 5,000 years ago they began to write about themselves and the world around them. The Mesopotamian was the first of the world's great civilizations.

For better and for worse, Mesopotamian civilization was built on an open, difficult-to-defend land. As a consequence Mesopotamian cities were occasionally razed by conquering armies and occasionally enriched by immigration. This went on for millennia. The Mesopotamians adapted. Despite the onslaught of invading armies and the legions of newcomers Mesopotamian culture persisted. Cuneiform, their method of writing, was in use for about 3,000 years. For well over 1,000 years they had the most advanced mathematics on the planet. They also had, as far as is known, the first written code of laws, the Hammurabi Code, which consisted of 282 laws that governed a wide variety of social and legal situations. These people fashioned one of the most creative and durable civilizations in history.

Eventually, however, Mesopotamian civilization fell behind other cultures. The last cuneiform records—records that contained information about astronomy—were made in the first century C.E. In time the culture of Mesopotamia was in large measure forgotten. It was not entirely forgotten, of course. The Hanging Gardens of Babylon were described by ancient travelers as one of the Seven Wonders of the Ancient World. There are a number of references to Babylon, one of the major Mesopotamian cities, in the Bible, and



Survey tablet. Written in cuneiform, this tablet records information about the dimensions of 11 fields as well as data about the amount of barley grown on each field. (The British Museum/Topham-HIP/The Image Works)

some of their ruins have always been visible to the interested traveler, but the extent of the accomplishments of the people of Mesopotamia is a fairly recent historical discovery. Unlike those of the Egyptians, who built monuments so massive and durable that they are impossible to overlook, and so serve as a constant reminder of Egyptian civilization, most traces of Mesopotamian architecture were wiped out or buried beneath the sands. Most of what we now know about Mesopotamia was unknown as late as 150 years ago.

Mesopotamian civilization was rediscovered in the 19th century as scholars began to decode the few examples of cuneiform writing that were then known. It was slow, difficult work, but the more they learned, the more interested they became. As archaeologists went out into the field to

uncover other Mesopotamian sites they unearthed enormous numbers of clay tablets, which were covered with cuneiform writing. Luckily for us, clay tablets proved to be an extremely durable medium, and tens of thousands of tablets were sometimes uncovered at single sites. The tablets address many aspects of life in this ancient civilization. Archaeologists found letters sent home from students to parents. There were legal records. There were stories and legends and poetry pressed into the clay. There were tables of

A MESOPOTAMIAN EDUCATION

We know quite a bit about Mesopotamian education because, among the many tablets detailing wars, business transactions, astronomical calculations, and government records, there are school books (or at least school tablets) and among these tablets there are a number of descriptions of a day in the life of a student. We know, for example, that the school day lasted from dawn to dusk. Students—practically all were boys—attended school for years. It was a difficult life. The following translation of a cuneiform text describes a day in the life of a schoolboy some 4,000 years ago.

I recited my tablet, ate my lunch, prepared my (new) tablet, wrote it, finished it; then my model tablets were brought to me; and in the afternoon my exercise tablets were brought to me. When school was dismissed, I went home, entered the house, and found my father sitting there. I explained my exercise-tablets to my father, recited my tablet to him, and he was delighted.

(Kramer, Samuel Noah. *The Sumerians, Their History, Culture, and Character*. Chicago: University of Chicago Press. Copyright 1963 by the University of Chicago. Used by permission)

Not all stories are this positive, but except for the accounts of caning, which was the preferred method of disciplining students, many of the stories sound remarkably modern.



[2] [3] [4] [5] [6] [7] [8] [9] Centimetres

Student tablet. Mesopotamian students learned how to write and do arithmetic on these small, bun-shaped clay tablets. (© Copyright The British Museum)

numbers, and astronomical computations, and lots of math problems. It was an intimate portrait of the first of the world's great civilizations. It is through these clay tablets that we now know so much about this ancient culture and its mathematics.

The Mesopotamian Number System

One of the main subjects taught in school was mathematics. The mathematics of the Mesopotamians was quite sophisticated. Becoming accustomed to it, however, takes a while because their number system was unique.

One of the characteristics that distinguishes Mesopotamian mathematics from all other systems of mathematics over the last 5,000 years is that the Mesopotamians used base 60 in their number system. The technical term for a base 60 system of notation is a *sexagesimal* system. Some remnants of this system survive today in our own number system. For example, we divide our minutes into 60 seconds and our hours into 60 minutes. (This gives 60×60 or 3,600 seconds in an hour.) To go back to the Malagasy method of counting soldiers with pebbles described in the previous chapter, counting in base 60 allows us to place up to 59 pebbles in the 1s pile. Each pebble in this pile represents an individual soldier. When the 60th pebble is placed in the 1s pile, it is wiped clean and a single pebble is placed in the 60s pile. Each pebble in the 60s pile represents 60 soldiers. When there are 60 pebbles in the 60s pile, the pile is wiped clean and a single pebble placed in the 3,600s pile ($3,600 = 60 \times 60$). The choice of base 60 is unusual, and there has been a lot of speculation about why the Mesopotamians chose to count in this way. It turns out that there are some advantages to counting in base 60—advantages that we explore later in this chapter.

The Mesopotamians did more than count aloud. They also developed a way to write these numbers. (Writing, of course, is a separate problem from counting.) Cuneiform, the written language of Mesopotamia, was produced by pressing the end of a stylus into wet clay. The illustration on page 16 shows some Mesopotamian number symbols.

To write numbers larger than 59 they developed a system of *positional numeration*, which enabled them to “recycle” the first 59 symbols into the next position. The symbols remain the same but the *meaning* of the symbols changes, depending upon the column in which they appear. The concept is simply a written version of the counting method used in Madagascar. When counting with stones, the number that the pebble represents depends on the pile in which the pebble is placed. Similarly in positional numeration the meaning of the symbol depends on the column in which it is written. As our system of positional numeration does, the Mesopotamian system allows the user to “build up” expressions for very large numbers without expanding the number of symbols required to make these numbers. The Mesopotamians used 59 symbols to write numbers in their system; we use 10 symbols. Not every culture that developed a method of writing numbers also developed a method of positional numeration. It is to their credit that the Mesopotamians were able to do both. The development of a written system of positional numeration by the Mesopotamians was a tremendous accomplishment. Their method is almost modern.

The difference in bases is not the fundamental difference between our system and that of the Mesopotamians, however. The essential difference between the two systems is revealed by the fact we use 10 symbols to express our numbers in base 10, whereas for a long time they used 59 symbols to express their numbers in base 60. We use exactly as many symbols as the base in which we count. They used one fewer symbol than the base in which they counted. In hindsight it is easy to see that they were one symbol short. They were missing the symbol for 0.

The need for a symbol to represent nothing is not immediately clear to most people. The Mesopotamians were no exception. Initially they seemed to give no thought at all to the matter. Instead they relied on the context in which the number appeared to convey a sense of its meaning. To use the Madagascar example again: The Mesopotamians ignored the empty pile and allowed the reader to guess which pile represented which power of 60. For example, two 1s in adjacent columns might represent one group of 60 plus 1 ($1 \times 60 + 1$). We would write this number as 61. Or the

same symbols might represent one group of 3,600 and one group of 60 ($1 \times 3,600 + 1 \times 60$)—the number that we would write as 3,660. (The 1s column was empty, but since they did not use 0 as a placeholder, there was nothing to mark its presence; in this situation the 1s column was absent.) Or two adjacent 1s might represent one group of 3,600 plus 1 ($1 \times 3,600 + 1$)—the number that we would write as 3,601. Presumably the difference between the number 3,660 and the number 61 is so large as to be clear to the reader from the context in which the symbols appear. The difference between 3,660 and 3,601 is not so apparent. Sometimes the Mesopotamians inserted a blank space between columns of symbols. This allowed them to distinguish between the numbers 3,601 and 3,660, for example. (Notice that 3,660 and 61 are still indistinguishable.) Leaving a space is clearly an improvement, but it is far from ideal since there is still a lot of ambiguity left in their notation.

By the time the Mesopotamian region was conquered by Alexander the Great in 331 B.C.E., the Mesopotamians had invented a symbol to represent *some* of the empty columns that

⋮	1	⋮⋮	2	⋮⋮⋮⋮	3	⋮⋮⋮	4
⋮⋮⋮	5	⋮⋮⋮⋮	6	⋮⋮⋮⋮⋮	7	⋮⋮⋮⋮⋮	8
⋮⋮⋮⋮⋮	9	↖	10	↖⋮	11	↖⋮⋮	12
↖⋮⋮⋮⋮	13	↖⋮⋮⋮	14	↖⋮⋮⋮⋮	15	↖⋮⋮⋮⋮⋮	16
↖⋮⋮⋮⋮⋮	17	↖⋮⋮⋮⋮⋮	18	↖⋮⋮⋮⋮⋮⋮	19	↖↖	20
↖↖↖	30	↖↖↖	40	↖↖↖	50	⋮	60

Cuneiform method of representing some of the numbers between 1 and 60

arise when writing numbers. If the empty column occurred between two nonempty columns, the Mesopotamians used two diagonal arrows to indicate the presence of the empty column. With this symbol the Mesopotamians were very close to finding a symbol equivalent to our 0, but they did not quite make the conceptual jump to a system of complete positional numeration. As far as is known the Mesopotamians did not use their symbol for an empty column when the empty column was the column farthest to the right—their 1s column. As a consequence this notation allowed them to distinguish easily between 3,601 and 61 but, again, not between the number 1 and the number 60, since in their system of positional notation both are written with their symbol for 1.

Of all the number systems of antiquity the Mesopotamian system most closely resembles a completely positional notation. Neither the Egyptians nor the Greeks nor the Romans developed a system of positional notation. Nor did these cultures develop notation as close to a positional system as did the Mesopotamians. The Mesopotamia system of notation was never true positional notation, however. Right to the end—for a period of about 3,000 years—they continued to depend on the context in which the symbol appeared to distinguish between certain numbers. The discovery of a symbol for 0—in fact the concept of 0—would have to wait centuries until Hindu mathematicians invented the symbol: the same symbol that we use today.

Advantages of Base 60 Notation

The Mesopotamians went quite a bit further than what we have described so far: They also developed a system for fractions that was similar to our decimal system, although they continued to use base 60 rather than base 10. In what we would call the 10ths column, which is the column immediately to the right of the decimal point, they had the $1/60$ column. In the position that we call the 100ths column, they had the $1/60^2$ ($= 1/3,600$) column. They included as many columns as they thought were necessary to maintain the accuracy of their calculations. The column immediately to the right of the $1/60^2$ column was the $1/60^3$ column

MESOPOTAMIAN MATHEMATICS HOMEWORK

This is a translation of a “problem text,” a clay tablet version of a textbook used for instructional purposes. We include it here because the numbers that appear in the problem are written in base 60. (The notation used in the problem is described in the main body of this chapter.) The problem itself uses a well-known mathematical result often attributed to the Greek mathematician Pythagoras: the Pythagorean theorem. Pythagoras, however, was born more than 1,000 years after this problem was written.

This text consists of the problem—which ends with the question, What is the breadth?—and an explanation of the algorithm used to solve the problem. The answer is identified as 0;10, and it concludes with the phrase “The method,” which was the standard closing for Mesopotamian mathematics problems.

The height is 0;40 (cubits),
 the diagonal 0;41 13 51 48.
 What is the breadth? You
 square 0;41 13 51 48. You
 will see 0;28 20. Square
 0;40, the height. You will see
 0;26 40. Take 0;26 40 from
 0;28 20. You will see 0;01 40. What is the square root? The
 square root is 0;10, the breadth. The method.

(Robson, Eleanor. Mesopotamian Mathematics, 2100–1600 BC, Technical Constants in Bureaucracy and Education. Oxford, United Kingdom. 1999. Reprinted by permission of Oxford University Press)

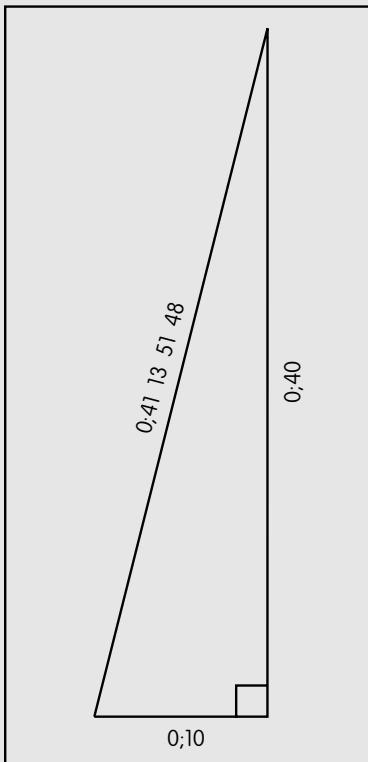


Diagram of a problem described in the text. A cubit is approximately 18 inches (45 cm) long. The measurements are in sexagesimal notation. The scale of the drawing is 6 in = 1 cubit (15 cm = 1 cubit).

($1/60^3 = 1/216,000$), and so on. It is in the writing of numbers smaller than 1 that we can get some insight into why the Mesopotamians may have preferred a sexagesimal system. A sexagesimal system is, in some ways, more convenient for calculation than is our base 10 system.

Before we examine the Mesopotamian method of writing fractions we introduce some convenient notation that all descriptions of Mesopotamian mathematics employ. We need to introduce a punctuation mark that will work as our decimal point does. Since the Mesopotamians used sexagesimal notation instead of decimal, we do not use a point. Instead we use a semicolon (;) to separate the 1s column from the $1/60$ column. (We use a semicolon so that we do not confuse base 10 numbers with numbers in base 60.) Furthermore we use a space to separate every column to the right of the semicolon. For example, the number that we might write as $1\frac{2}{60}$ is written as the sexagesimal number 1;02. (We mark the $1/60$ column with two decimal digits—hence the 02 instead of simply the number 2.)

To see the advantage of computing with sexagesimal notation we need only consider a common fraction such as $1\frac{2}{3}$. In our decimal notation the number $1\frac{2}{3}$ is written 1.66666. . . . The *decimal* expression does not terminate; the 6s go on forever. The point at which you stop writing 6s and “roundoff” is the step in which you introduce an inaccuracy. But $2/3$ can also be written as $40/60$. So in base 60 the number $1\frac{2}{3}$ is written as 1;40. In sexagesimal notation the $2/3$ is much easier to write and there is no round-off error. Here is one more example: In base 10 the number $2\frac{1}{12}$ is written as 2.083333. . . . The 3s carry on forever. But because $1/12$ can also be written as $5/60$, in base 60 the number $2\frac{1}{12}$ is written as 2;05. The sexagesimal expression is much simpler than the decimal expression of the same number and there is no round-off error.

It turns out that many common fractions are written much more simply in base 60—once you get used to it!—than they are in our own decimal system. This may explain why the Mesopotamians chose the sexagesimal system. It made their computations easier to perform accurately. We should not

forget, however, that the lack of a 0 meant that the system that they employed for writing numbers with fractional parts had the same ambiguities that are to be found in their system for writing counting numbers.

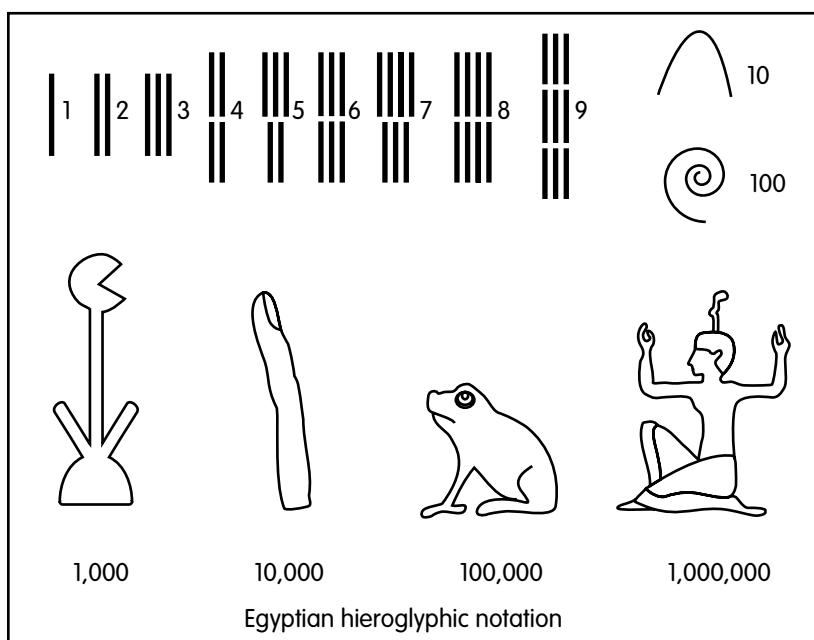
The translation of ancient Mesopotamian texts and the search for their mathematical and cultural meanings remain active areas of scholarship. Part of the difficulty in understanding cuneiform texts is that the Mesopotamians had a very different approach to mathematics and a different understanding of the role of mathematics in society from ours. Scholars are still trying to sort out the differences. We know that they were interested in a kind of algebraic geometry, in which geometric ideas were expressed in a sort of protoalgebra. Sometimes this type of math is called numerical algebra. Whatever we call it, we know that they put their knowledge to work when they divided estates, predicted astronomical phenomena, and calculated budgets for construction projects.

These applications in themselves do not explain the Mesopotamians' intense interest in mathematics. They computed extremely accurate approximations to certain numbers—the square root of 2 is a well-known example. The approximation, which is carried out to the 1,000,000th decimal place, is much more accurate than is required for practical work. Why did they bother? Did they just enjoy a mathematical challenge, or did the work have a deeper meaning to them? These are questions that are open for debate. Much current scholarship centers less on what the texts say than on what the texts tell us about the role of mathematics in this earliest of civilizations.

The Egyptian Number System

Less information is available about Egyptian mathematics than about Mesopotamian mathematics, but one point is clear: Egyptian mathematics was not as advanced as Mesopotamian mathematics. One reason that less information is available about Egyptian mathematics is that the Egyptians wrote on papyrus whereas the Mesopotamians wrote on clay. Clay lasts longer.

Much of what we do know about Egyptian mathematics is taken from two scrolls. The more famous of the two is called by one of two names, the Rhind papyrus or the Ahmes papyrus. Alexander Henry Rhind purchased the largest part of the papyrus while on a trip to Egypt in the 19th century. (Other parts were later located in the collection of the New-York Historical Society.) Ahmes was the scribe who originally produced the text, which was written about 3,700 years ago. We refer to the text as the Ahmes papyrus, since, of the two, Ahmes did more work than Rhind. The other important text is called the Moscow papyrus. Less has been written about the Moscow papyrus, in part because it is a smaller work. The Ahmes papyrus was originally about 18 feet (5.5 m) long, and 13 inches (33 cm) wide. The Moscow papyrus is also about 18 feet (5.5 m) long, but it is only about 3 inches (8 cm) wide. Both scrolls are collections of mathematics problems.



Egyptian hieroglyphic notation. Whole numbers were represented by combining the appropriate multiple of each symbol of a power of 10.

Although the Egyptians counted in base 10 they had more than one way of writing numbers. The better-known form is hieroglyphic notation. In hieroglyphic notation there is no concept of positional numeration at all. Each power of 10 has its own symbol. For example, the number 1 is represented by a simple stroke, much as our own is. The number 10 is represented by a horse-shoelike mark. The number 100 is represented by a curved line that is very similar to the top of a fiddlehead fern. To represent any counting number between 1 and 999 one simply wrote the required number of strokes, horseshoes, and fiddleheads. For example, 254 was written as two fiddleheads, five horseshoes, and four strokes. As a consequence in hieroglyphics the Egyptians could count as high as they had symbols for powers of 10, but no higher. If they needed to count higher they had to devise new symbols. Unlike the Mesopotamians, the Egyptians could not use hieroglyphics to write very large numbers without inventing new notation.

The writing in the Ahmes papyrus, however, is in hieratic script. Hieratic script was generally written in ink on papyrus, and it is a good deal more compact than hieroglyphics. That is its advantage. Its disadvantage is that there are separate symbols for 1, 2, 3, 4, 5, 6, 7, 8, and 9 as well as 10, 20, 30, . . . , 90; as well as 100, 200, 300, 400, . . . , 900; as well as 1,000, 2,000, 3,000, . . . , 9,000. To write the number 456 in hieratic script only three symbols are necessary, one symbol for 400, one symbol for 50, and one symbol for 6. To write the same number in hieroglyphics, 15 symbols are necessary: four symbols for 100 (four fiddleheads), five symbols for 10 (five horseshoes), and six symbols for the number 6 (six strokes). There was no concept of positional numeration in the hieratic system, either. If the order of the numerals was interchanged in hieratic notation the value of the number remained the same.

For writing fractional parts of numbers the Egyptians had an unusual system of notation. This system influenced the development of the concept of number in nearby cultures, including that of the Greeks. The Egyptians used only fractions of a very special form. We would write this type of fraction with a counting

1		10		100		1000	
2		20		200		2000	
3		30		300		3000	
4		40		400		4000	
5		50		500		5000	
6		60		600		6000	
7		70		700		7000	
8		80		800		8000	
9		90		900		9000	

Hieratic numerals

Hieratic numerals. This system used a large number of specialized number symbols. Its advantage is that the notation is more compact than hieroglyphic notation.

A PROBLEM FROM THE AHMES PAPYRUS

The Ahmes papyrus contains many problems that involve the equal distribution of bread and beer. Problems involving the distribution of bread and beer are not the only types of problems that appear in the Ahmes papyrus, however. There are, for example, geometry problems and problems that are simply computations without any additional motivation. Some of these problems would be difficult for many well-educated laypersons to solve today, but not all of them would. Today some would be considered trivial. At the time the Ahmes papyrus was written, however, the work involved in obtaining a solution to even a simple problem could be extensive because of the primitive notation and primitive methods of computation that the Egyptians used. Their math was primitive when compared with ours, of course, but it was also primitive compared with that of their contemporaries the Mesopotamians. It is clear from the problem included here that the system of notation used by the ancient Egyptians could make solving even simple problems difficult. The following bread problem is an example of an easy problem with a laborious solution. It is the fifth problem in the Ahmes papyrus, one of the simplest. The method of solution is described in the paragraph following the translation. (It is standard to identify those lines that are used in the computation of the final answer with a back slash.)

Problem 5

Divide 8 loaves (of bread) among 10 men.

Each man receives $2/3\ 1/10\ 1/30$.

Proof. Multiply $2/3\ 1/10\ 1/30$ by 10; the result is 8.

number in the denominator and a 1 in the numerator. We call fractions with a whole number in the denominator and a 1 in the numerator *unit fractions*. The ancient Egyptians wrote these fractions by writing the counting number that appeared in the denominator and then drawing a dot or an oval above it. Because they considered only unit fractions, writing the denominator (with a dot) was enough to identify the fraction. For example, $1/2$, $1/5$, and $1/96$ are all unit fractions. By contrast the number

Do it thus:	1	2/3	1/10	1/30
	\2	1 1/2	1/10	
	4	3 1/5		
	\8	6 1/3	1/15	

Total 8 loaves, which is correct.

(Reproduced with permission from the Rhind Mathematical Papyrus, Free Translation, and Commentary, 1929 by the National Council of Teachers of Mathematics)

Each entry in the right column represents what we would think of as a single number. (Keep in mind that the Egyptians used only the fraction 2/3 and unit fractions. So instead of writing 8/10 in the top row, Ahmes writes this number as 2/3 1/10 1/30, because $2/3 + 1/10 + 1/30 = 8/10$. Each entry in the right column is a fraction or mixed number. To express it in a form that we would easily recognize just add all the fractions together.) Each entry in the right column is 8/10 the size of the corresponding number in the left column. Each of the entries in each row is twice as big as the corresponding entry in the row immediately above it.

For the Egyptians multiplication meant multiplication by 2. To multiply any two whole numbers, they just performed a series of doublings and kept track of each step. They solved the problem by adding some of the intermediate steps—here marked with a back slash (\)—to get their final answer. Here is his method:

To get 10 men, he adds 2 and 8 in the left column. Next he adds the corresponding numbers in the right column—that is 1 1/2 1/10 plus 6 1/3 1/15—to get 8, for the 8 loaves of bread. This proves that the first line is the correct answer: 1 man should receive 2/3 1/10 1/30 or 8/10 of a loaf of bread. It is a lot of work to solve such an easy problem.

3/4 is not a unit fraction, because it has a 3 in the numerator. (For the Egyptians the only exception to this rule was the fraction 2/3, for which they had a separate symbol.) Because they used only unit fractions, they had to write any other fraction as a sum of unit fractions. For example, the number we write as 3/4, they wrote as 1/2, 1/4 ($1/2 + 1/4 = 3/4$). Another example: The number that we would write as 21/30 they might write as 1/3, 1/5, 1/6, or as 1/2, 1/5, 1/30. Either set of fractions adds up to

the number we know as $21/30$. (There is generally more than one way to write a fraction as a sum of unit fractions.) Notice, too, that they did not put addition signs between the fractions. They maintained this system of writing fractions, and they computed with this somewhat awkward system, for thousands of years. From our viewpoint it is difficult to see any advantages to this type of notation.

The Mayan Number System

The Mayan civilization was centered in the area around the Yucatan Peninsula in what is now Mexico. Less is known about Mayan culture than is known of Mesopotamian or Egyptian culture because fewer records survive. We know that the Mayan written language used hieroglyphics, and we know that they had libraries with thousands of books. Unfortunately after the Spanish conquered the Mayans during the first part of the 16th century they set out to burn all Mayan books. They nearly succeeded. Today there are four more-or-less complete books left, the Dresden Codex, the Perez Codex, the Codex Tro-Cortesianus, and the Grolier Codex. None of these codices is in very good condition.

Mayan books were not constructed as the books that we know are. They are long sheets of paperlike material that are folded back and forth as a fan is. By flipping the folds the reader moves from page to page. The text with the most mathematics on it, the Dresden Codex, was damaged during the bombing of Dresden, Germany, late in World War II.

We can obtain a little more firsthand information about the Mayans from Spanish accounts written during and just after the conquest, although these accounts are sometimes more sensational than informative. In any case the Classic period, the time of greatest Mayan cultural innovation, ended centuries before the arrival of the Spanish. The Classic period lasted from about C.E. 300 to C.E. 900, when the Mayans, during a long period of extreme drought, abandoned many of their largest cities.

Despite these difficulties the Mayan number system is well worth investigation because it is unique and in certain aspects it

is very sophisticated. Much of the Mayans' motivation for developing a sophisticated system of mathematics was their interest in astronomy and calendars. They worked hard to predict astronomical phenomena such as eclipses and to maintain an accurate calendar. The Mayan calendar was quite complicated and was at least as accurate as any calendar in the world at the time it was created.

The Mayan system of numeration is often described as base 20. This is not quite correct. The Mayans maintained two systems of numeration. One had religious significance and used a combination of multiples of 20 and multiples of 360. Although this system was important to the Mayans, Spanish accounts make clear that there was a second system of numeration in common usage. This "common system" *was* a base 20 system. This is the system on which we concentrate. Referring, again, to the story in the first chapter about the Malagasy system of counting, the Mayans would have placed up to 19 pebbles in the 1s pile. When the 20th pebble was to be added, they would have wiped the 1s pile clean and added a single pebble to the 20s pile. One stone in the first pile would represent a single unit, one stone in the second pile would represent 20 units,



A page from the Dresden Codex, one of the few Mayan books to survive to modern times (ARPL/Topham-HIP The Image Works)

a single stone in the third pile would represent 400 units— $20 \times 20 = 400$ —and one stone in the fourth pile would represent a group of 8,000 ($8,000 = 20 \times 20 \times 20$). Just as our system is ordered around powers of 10, the common Mayan system of numeration was organized around powers of 20. A base 20 system of numeration is a *vigesimal* system.

0	1	2	3	4
5	•	••	•••	••••
10	•	••	•••	••••
15	•	••	•••	••••
20	•	••	•••	••••
25	•	••	•••	••••

Mayan positional number system

17	+	22	=	39

The Mayan system of vigesimal notation. Notice the use of the symbol for 0 in the number 20. This method of notation was well suited for quick whole number calculations.

The Mayans expressed their numbers in their secondary or common system via true positional notation. In addition to their symbols for the digits 1 through 19 they had a symbol for 0. They were one of the very few cultures in history that could see far enough into mathematics to recognize the need for, and the utility of, a symbol for nothing. The symbols that the Mayans used to represent their numbers are shown in the illustration. The symbol for 0 is represented by an oblong figure representing a shell. Each of the other 19 symbols in their system was represented by a combination of dots and dashes. Each dash represents 5. Each dot represents 1. The symbols were arranged vertically so that addition was a simple matter of merging the symbols in the corresponding positions and then, if necessary, "carrying" a digit into the next column. The Mayans apparently were quite skilled at this procedure and could perform extensive arithmetic manipulations with a few handfuls of beans. It was this facility with numbers that so impressed the Spanish colonists (see the illustration). This is a very convenient system for whole number arithmetic.

The Mayans spent a great deal of time developing their mathematics. Their astronomical calculations were surprisingly accurate. For example they calculated that 149 lunar months lasted 4,400 days. This is how *they* expressed their estimate of a lunar month. If we divide 4,400 days by 149 lunar months we get a more modern expression for the length of a lunar month. The difference between the Mayan estimate for a lunar month and the actual value is about one minute, an extraordinary accomplishment, especially when one considers that they had no modern measurement devices and their mathematics *did not use fractions*. There is no evidence that the Mayans incorporated any sort of fractional notation in their vigesimal system. Since they had already solved all of the hard problems with respect to developing a complete system of positional notation it is difficult to see why they did not go further and develop a system that included vigesimal fractions. Apparently, however, what they invented was adequate for the computations that they wanted to perform.

The Chinese Number System

Chinese mathematics may be as ancient as that of Mesopotamia. Early records are scarce. We know that a system of writing was in use during the 17th century B.C.E. and that this early writing system had some similarities to contemporary Chinese writing. Presumably they produced a number of mathematics texts at this time, but none survives. In 221 B.C.E. China was united under the rule of Qin Shi Huang (also known as Shih Huang-ti), the first emperor of China. He established a strong central government that undertook many large-scale construction projects. Under Qin Shi Huang's rule the Chinese built an extensive system of roads and fortified the Great Wall. They also instituted a uniform currency system. Despite these achievements the first emperor ruled harshly. He was a book burner: He ordered the destruction of all books. This is why our knowledge of Chinese mathematics largely begins after the end of the reign of Shi Huang.

We do know that two systems of numeration were in use in ancient China. Both used base 10. One system used a symbol for each number from 1 to 10 and then a separate symbol for each power of 10 thereafter. To write a number one simply wrote one of the numerals (from 1 to 9) followed by the power of 10 the number represented. For example the number 321 would be written with the numeral for 3 followed by the symbol for 100 followed by the numeral for 2 followed by the symbol for 10 and finally the numeral for 1. This is not quite positional notation but it is close.

The other system, called rod numerals, used only a set of straight lines to represent any counting number. The rod numerals are shown in the illustration. In this system there are separate symbols for the numbers 1 through 10 and for multiples of 10 up to 90: 20, 30, 40, . . . , 90. One can represent any counting number with these 18 symbols. Here is how: To represent the number 6,573, for example, write the symbol for 3, then the symbol for 70, followed by the symbol for 5, and finally the symbol for 60. (Alternating symbols in this way prevented ambiguity about where one symbol left off and the next began.) Where we would write a 0, the ancient Chinese left a blank space. This is almost a system of positional

	1	2	3	4	5	6	7	8	9
Units						T	T	TT	TTT
Hundreds									
Ten thousands									
Tens									
Thousands	—	=	≡	≡≡	≡≡≡	⊥	⊥	⊥⊥	⊥⊥⊥

Chinese rod numerals. By alternating vertical and horizontal rods and by leaving a blank space for 0, the Chinese invented a system of notation that made simple calculations quick and easy.

notation. Rod numerals were in use for a thousand years before Chinese mathematicians began using a symbol for 0. The symbol for 0 that they used is identical to the one we use today.

What is interesting about rod numerals is that they did not just resemble rods; in many cases the rod numerals were actually made with rods. Individuals who were required to perform arithmetic in the course of their job often carried a bag full of rods. Sometimes the rods were colored differently for positive and negative numbers. The rods were used as an aid to calculation. The numbers were constructed and then manipulated by adding and subtracting rods. A lot has been written about the speed and accuracy that these workers could attain with the help of their rods.

One of the earliest surviving Chinese mathematical texts is called *The Nine Chapters on the Mathematical Art* (also translated as *Arithmetic in Nine Sections*). The *Nine Chapters*, like the Ahmes papyrus, is a series of math problems on a variety of topics. There are problems in algebra and geometry as well as arithmetic. Not all the problems are correctly solved. Nevertheless the problems are advanced for the time and so are many of the solutions. There are problems involving fractions as well as whole numbers, and some of the methods of solution that are employed in the text are similar to the methods that we use for doing arithmetic with fractions today. The book also contains fractions expressed in a kind

A PROBLEM FROM THE NINE CHAPTERS

The Nine Chapters on the Mathematical Art was described by a Chinese writer living in the third century C.E. as a text that survived the period of book burnings ordered by the emperor Qin Shi Huang and was revised and expanded upon by more than one author. The revised edition first appeared early in the Han dynasty, which lasted from 206 B.C.E. until C.E. 220. The problems in the *Nine Chapters* still pose a healthy challenge to many junior high and high school students. The analyses that follow the problems are not always easy for the modern reader to follow because of the lack of algebraic symbolism. They are just too wordy, but many of the problems (and a few of the analyses) have a modern ring to them; they could be from a contemporary textbook. Here are a very readable math problem and its accompanying solution from the Han dynasty. (Notice the use of the Pythagorean theorem in the solution of the problem.)

Under a tree 20 feet high and three feet in circumference, there grows an arrowroot vine, which winds seven times (around) the stem of the tree and just reaches the top. How long is the vine?

Rule. Take 7×3 for the second side (of a right triangle) and the tree's height for the first side. Then the hypotenuse is the length of the vine.

(Mikami, Yoshio. *The Development of Mathematics in China and Japan*. New York: Chelsea Publishing, 1913)

of decimal notation as well as problems that represent a single fraction as a sum of unit fractions, in the manner of the ancient Egyptians. Some—though by no means all—of the problems in *The Nine Chapters on the Mathematical Art* have a very modern ring to them.

The effect of the *Nine Chapters* on subsequent generations of Chinese mathematicians was profound. Mathematicians wrote commentaries on the *Nine Chapters* for centuries after the book first appeared early in the Han dynasty. (The Han dynasty was established in 206 B.C.E., five years after the death of the first Chinese emperor, Qin Shi Huang.)

There are many other ancient cultures whose system of numeration we could examine. The Greeks, the Romans, the Etruscans, the Japanese, and others developed their own individual systems of numeration. The Greeks, for example, are famous for the relative sophistication of their geometry, but their system of numeration was no more or less advanced than their neighbors'. The cultures that we have considered so far give a good sense of the mathematical diversity that was present many centuries ago. Each culture developed a unique system of numeration. They did not just differ in the way they wrote numbers or even in the bases that they used to express their ideas about numbers. There was also variation in their concept of what a number is. The Mesopotamians, for example, developed a very precise method of expressing arbitrary fractions. The Egyptians considered only unit fractions and the Mayans used no fractions at all. Each system had qualities that, in hindsight, facilitated or inhibited calculation, but each system, apparently, satisfied the needs of the users, whether they used the system to compute taxes or the dates of lunar eclipses.

3

OUR PLACE VALUE NUMBER SYSTEM

The cultures of ancient China, Egypt, and Mesopotamia relied on experts to perform even simple calculations. They had no choice. Their systems of notation were so awkward that only the most practiced of mathematicians could perform arithmetic manipulations of moderate difficulty. (See, for example, the problem of dividing eight loaves of bread among 10 men in the Ahmes papyrus.) By contrast many of these previously difficult computational problems are now easy enough to be solved by schoolchildren.

The two great differences between arithmetic now and arithmetic in antiquity lie in the introduction of 0 and the truly positional place value notation that the concept of 0 made possible. The “invention of zero,” although it seems almost too simple to matter, was a tremendous mathematical innovation. Many problems that were previously beyond reach became easily solvable. The frontiers of computation were greatly extended, and once a completely decimal system was adopted, computational techniques and science in general surged forward.

There is not much diversity left in the world with respect to systems of numeration used for hand calculations. Cultures all over the world now use base 10 to count, and it is a near-universal practice to employ 10 symbols—in the West those symbols are 0, 1, 2, . . . , 9—in a system of positional numeration. This is a testament to the superiority of the present system to what preceded it. Despite its importance to mathematics, not much is known about the way our current system of notation was



Published in 1503, this woodcut represents two competing methods of performing computations. The figure on the left is using the “new” Arabic numerals. The figure on the right is using a counting board, a device that is similar in concept to an abacus. The figure in the center represents arithmetic. (Library of Congress, Prints and Photographs Division)

established. Many conflicting reports exist. We do know that books and letters describing the new system were written during the seventh century C.E., a time of interaction among many of the most mathematically advanced cultures of the age. Ideas about

mathematics and religion were exchanged between India and China to the east and between India and the emerging Islamic culture to the west, but the source of the individual contributions is often unclear. Indian mathematicians are usually given credit for devising our modern system of numeration, but it is not clear what they invented and what they borrowed. We have seen, for example, that centuries before the Indians, the Mesopotamians had an almost-complete system of positional notation, and that the Mesopotamians had a symbol that was used in a way that was somewhat similar, but not identical, to our own 0. The Mesopotamian system was base 60, however, not base 10. Others give the Egyptians credit for inventing a symbol for 0. Others give credit to the Greeks. One fact seems clear: The Hindu mathematicians were the first to join these ideas in one system of numeration.

It is customary to refer to mathematics that originated in the subcontinent of India as Hindu mathematics; unfortunately, there is not at present a more descriptive term. India is huge and culturally and religiously diverse. The Buddha was roughly a contemporary of Pythagoras (ca. 580 B.C.E.–ca. 500 B.C.E.). The history of the area is complicated and ancient. The mathematics that has emerged from this area is not uniform, and whether mathematicians in one part of the region were always aware of the work of mathematicians in another is not entirely clear. There are some common threads in the mathematics that originated in this area, however. One is a preference for algebra. Though the ancient mathematicians of present-day India almost certainly had some awareness of the geometry of the Greeks, they showed little interest in it. They were interested in algebra and trigonometry. (The sine function, the first of our six trigonometric functions, originated in India.)

The expositions of Hindu mathematics that have survived often read more like poetry than mathematics, and most of the math texts were, in fact, written in verse. This may have made them more fun to read, but it often caused the author to leave out the less poetic aspects of the subject matter, such as rational justification for certain formulas or techniques. Today no account of

Hindu mathematics omits the assertion that it was in what is now India that our modern system of numeration was developed.

It is not clear when the mathematicians of this region began using what has become our system of numeration. We do know that in C.E. 662 a Syrian bishop named Severus Sebokt wrote a glowing description of the calculating system employed by Hindu mathematicians, a system that used *nine* symbols. This indicates two points: (1) They were using base 10 by C.E. 662, and (2) the account of Hindu arithmetic that had fallen into Sebokt's hands did not contain the symbol for 0. (Otherwise he would have asserted that they computed with 10 symbols.) Further we know that there are Indian documents from the late ninth century that contain the symbol for 0, so sometime between C.E. 662 and C.E. 900 the modern system of numeration was born. Hindu mathematicians did not extend their system of positional notation to decimals, preferring to use common fractions instead. (Common fractions are numbers that are expressed in the form a/b where a and b are whole numbers and b is not 0.)

It would be nice to report that the advantages of this new system of numeration were immediately recognized. This, however, was not always the case. Islamic culture was fairly quick to adapt. In the latter part of the eighth century a Hindu astronomical text, *Siddhānta*, was translated into Arabic. It was described in detail in the writings of the highly influential Islamic mathematician Mohammed ibn-Mūsā al-Khwārizmī (ca. 780–ca. 850). Soon the Hindu system of numeration was adopted throughout the Islamic world. In Europe, however, the process of change was slower.

There was quite a bit of resistance to the adoption of the Hindu ideas in Europe. The first known mathematical treatise to introduce the Hindu system of numeration to Europe and to argue for its adoption was written by Leonardo Fibonacci, also known as Leonardo of Pisa. His family was from Pisa, Italy, but they maintained a business in northern Africa. Leonardo was educated by a Moor—the Moors, who hailed largely from North Africa, conquered Spain in the eighth century—and as an adult Leonardo traveled throughout northern Africa and Syria. While there he was exposed to several systems of numeration. Some say that he

EXPLAINING THE NEW SYSTEM

We are so accustomed to our system of numeration that imagining it as an innovation that had to be taught to adults familiar with another system of numeration is hard. That is why it is so interesting to read an early European account of the Hindu system. The following is an excerpt of a 15th-century English translation of a 13th-century Latin text written by John Holywood (ca. 1200–1256), also known as John of Halifax, also known as Sacrobosco. The title of the text is *The Art of Nombryng*, and in it he describes what was then the new system of Hindu numeration. As you read it, notice the points that he believes worth emphasizing.

- He emphasizes that every number can be represented as a combination of 10 symbols.
- He notes that the numerals 1 through 9 are represented by individual cursivelike symbols as opposed to the more complicated systems developed in Egyptian or Mesopotamian numeration: “All digits shall be written with one figure alone.”
- He is quite clear on the significance of the number 0.
- He stresses the economy of notation offered by the new system: “each number from one hundred to one thousand, which is excluded, shall be represented with .3. figures.”
- Finally, notice that he sets off each number but *not each digit* from the rest of the text with a pair of periods.

learned of the Hindu system from al-Khwārizmī’s book. Others assert that he learned it from a commentary on al-Khwārizmī’s book written by another Islamic mathematician. In any case there seems little doubt that it was from Islamic mathematicians that Leonardo learned of the Hindu system of numeration, and it was through Leonardo of Pisa that the Hindu system of numeration was introduced into Europe. In his best-known book, *Liber Abaci*, he describes the Hindu method of numeration and advocates its use. Interestingly he uses the Hindu method only for whole numbers. When dealing with fractional parts of a whole number he sometimes retains the ancient Egyptian habit of confining himself

It is a fine piece of mathematical writing from medieval Europe.

Understand that there are .9. characters or figures that represent the .9. digits; they are .9.8.7.6.5.4.3.2.1. The .10. (*tenth character*) is called a cipher. It gives significance to the others; for without a cipher a pure article could not be written. Thus with these nine characters and the cipher all numbers may be written. Note well that all digits shall be written with one figure alone, and every article with a cipher, for every article is named for one of the digits as .10. for .1., .20. for .2. and so for the others. All digits shall be set in the first place and all articles in the second. Also all numbers from .10. to .100., which is excluded, must be written with .2. figures. And if it be an article, by a cipher first, and the figure written towards the left, that signifies the digit by which the article is named; if it be a compound number, first write the digit that is a part of the compound then write at the left side the article. All numbers that are from one hundred to one thousand, which is excluded, shall be written with .3. figures, and all numbers from one thousand to ten thousand must be written by .4. figures; and so forth. And understand that every figure in the first place signifies a digit, in the second place .10. times his digit, in the third place one hundred times, etc.

(Reprinted with permission from Mathematics Teacher, Vol. 35, copyright 1942 by the National Council of Teachers of Mathematics. All rights reserved)

to unit fractions. Other times he expresses himself in the sexagesimal fractions of the Mesopotamians. Leonardo Fibonacci had one foot in the future and one in the past.

The obvious question is, Why did Europeans take centuries to adopt the Hindu system of numeration? At least part of the answer lies in the fact that many Europeans recognized the Hindu system of numeration as what it was: a powerful tool for computation. There was, however, already a powerful computing device firmly entrenched in European society, the abacus. In fact those who supported the use of Hindu numeration were called algorists and those who opposed it were called abacists. The abacists were

traditionalists who fiercely opposed the introduction of the new system. Both the Hindu system of notation and the abacus were perceived as being tools for quick and accurate computation. It may be difficult for us to see how positional numeration and the abacus could be perceived as mutually exclusive, but they were. The argument between the two groups went on for centuries.

The use of the Hindu system of numeration (for whole numbers) slowly became common practice throughout Europe, but until about 400 years ago fractions were still written as common fractions or in sexagesimal (base 60) notation. (If the continued use of base 60 sounds strange, it should not. We still measure time in a mixed system. We use base 60 for minutes and seconds, and we measure longitude and latitude in base 60 as well. The sexagesimal system of the Mesopotamians is one of the most enduring of all mathematical innovations.) By the late 16th century, however, there was great interest in developing more accurate, larger-scale computations for both scientific and commercial purposes. The computations were carried out with the help of tables of numbers that were used in much the same way that we would use hand calculators. Bankers, surveyors, scientists, and others who worked with numbers depended on the tables to aid them in their calculations. The entries in the tables were usually recorded by using the Hindu system of numeration for the whole number and the sexagesimal system for the fractional part. It was exhausting to compile the tables and, given the mixing of decimal and sexagesimal systems, it must have been exhausting to use them.

It was during this time that we can begin to find decimal fractions appearing in the work of several mathematicians. No one person invented decimal fractions, but three people, François Viète, Simon Stevin, and John Napier, were particularly instrumental in promoting the acceptance of the new system. A complete base 10 system, including decimal fractions, made a huge difference in the science of computation.

The French mathematician François Viète (1540–1603) has priority. He was the first to present a reasoned argument against the continued use of the sexagesimal system for representing fractions and for the adoption of a decimal system in its place. He present-

ed this argument in a supplement to his book *Canon Mathematicus* in 1579. In this work he tries to follow his own advice and uses several methods, more or less similar to our own, for representing decimal fractions. Where we would write 3.14, Viète sometimes wrote 3|14, using a vertical line where we would use a decimal point. Elsewhere he wrote 3^{14} or even 3^{14} . He was searching for the best method of representing decimals, but his notation was in every case so similar to our own that we would have no problem following any of his notations.

The Dutch mathematician and engineer Simon Stevin (1548–1620) is generally given credit for doing the most to establish the decimal system of notation throughout Europe. Many writers give him all the credit. Here is why: Stevin wrote a small but very influential book describing the new system of notation. First published in 1585, *The Tenth* was not aimed at fellow mathematicians. Instead Stevin wrote the book—it was really more of a pamphlet—for anyone whose job required measurement and computation. The introduction to the book is very explicit about the audience. Stevin states that surveyors, merchants, and people who work with tapestries, among others, all belong to his intended audience. In *The Tenth* he describes how one can use decimal fractions in addition, subtraction, multiplication, and division. He proves that his method of numeration, when used to do arithmetic, gives the correct answers, and, finally, he gives examples of how to use decimal notation when doing addition, subtraction, multiplication, and division. Though the book is short, it is very thorough. And because even very common fractions often yield “nonterminating” decimals—for example, $1/3$ in decimal notation is $0.333333\dots$ —Stevin even briefly discusses the effect of “round-off error,” which occurs when we write a terminating fraction in place of a nonterminating fraction. For example, we might write 0.33 in place of $0.333\dots$ and this introduces an error into the computation. Stevin assures the reader that one need only retain enough digits to get the desired accuracy. The ability to compute the thousandth part of an ounce in a business transaction, he says, is neither necessary nor desirable. He tells us that approximations can be more useful than perfection. His point of view is pragmatic

and modern. (As we see in chapter 4, we follow Stevin's maxims whenever we use a calculator, whether we know it or not.)

The effect of the book was profound and immediate. It was quickly translated into several languages and readers took Stevin's advice to heart. After giving the cause of decimal fractions a big boost Stevin went further. He argued for the adoption of not just a system of decimal fractions but also decimal systems of weights and measures and money. All of this has happened. The metric system has been adopted by almost every country on Earth (the United States is a notable exception), and metric measurements are the standard for all scientists, including those working in the United States. Even though the United States still uses a nondecimal system of weights and measures, like that of most of the rest of the world, its monetary system is decimalized.

Interestingly, though, of the three mathematicians we mention here, Stevin's notation for decimal fractions was the most awkward

and the least like our own. Unlike Viète's notation, Stevin's notation is not nearly so easy for the modern reader to follow. In order to keep track of which is the 10ths column, the 100ths, and so on. Stevin placed the symbol 0, a 0 in a circle, above the 1s column of a number. He placed the symbol 1 over the 10ths column. The symbol 2 was placed over the 100ths column. The 3 was placed over the 1,000ths column, and so on. See the illustration for a comparison of some of these early notations.

The Scottish mathematician John Napier (1550–1617) was

Viète: $\begin{array}{r} 2\ 15 \\ + 3\ 14 \\ \hline 5\ 29 \end{array}$ and $\begin{array}{r} 2 15 \\ + 3 14 \\ \hline 5 29 \end{array}$	Stevin: $\begin{array}{r} ①\ ②\ ③ \\ 2\ 1\ 5 \\ + 3\ 1\ 4 \\ \hline 5\ 2\ 9 \end{array}$
Napier: $\begin{array}{r} 2.15 \\ + 3.14 \\ \hline 5.29 \end{array}$	

The equation $2.15 + 3.14 = 5.29$ in the notations of Viète (two different methods), Stevin, and Napier

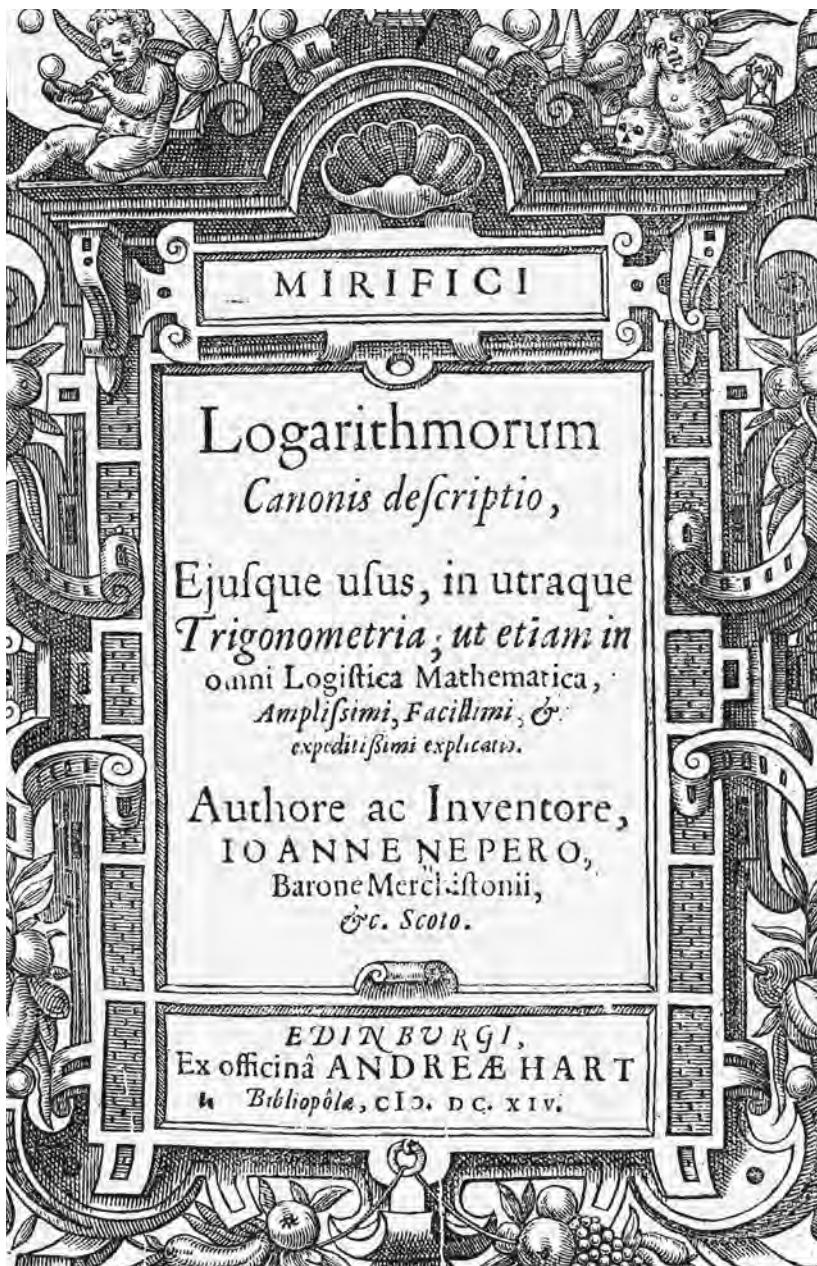
the third person who was influential in the adoption of our current system of decimal notation. Napier was very interested in problems associated with computation. He recognized early in life that one of the great barriers to scientific progress in his time was the difficulty that scientists encountered whenever they performed calculations. He set out to find ways to facilitate computation. (Remember: Most scientists were still making their calculations with a system that used base 10 for integers and base 60 for fractions.) In a variety of ways Napier permanently changed the way calculations were performed. He discovered a way to facilitate simple multiplication through the use of a set of rods that had sequences of numbers on them. These are called Napier's rods or Napier's bones. The rods are sometimes used in math classes today to broaden students' appreciation of arithmetic.

More importantly for us, Napier found a way to facilitate very large-scale computations by inventing logarithms. Today we understand logarithms as functions. Logarithms are often studied in calculus courses, though in Napier's time calculus had not yet been invented, so his understanding of logarithms was somewhat different from ours. Suffice it to say that Napier communicated his results in the form of tables of numbers. His work was among the first to express numbers by using a decimal point between the 1s position and the tenths position.

Because Napier's discoveries were very useful, his text proved very popular. Napier's book was the ideal advertisement for his complete system of base 10 notation including decimal fractions. Napier preferred to use a decimal point instead of the notations that had been used by Viète and Stevin, and the success of his book led to the widespread adoption of the decimal point.

Although the decimal notation is not fully standardized—today some countries use a *decimal comma* where others use a decimal point—the system of numeration that we use today has passed to us largely unchanged from Napier's 1619 text on logarithms, *Mirifici Logarithmorum Canonis Constructio*.

One more innovation in numeration during this period occurred after the time of John Napier and has no relation to decimal fractions. The innovator was Gottfried Wilhelm Leibniz (1646–1716),



John Napier's Mirifice Logarithmorum Canonis Descriptio, the book that popularized the use of the decimal point (Library of Congress, Prints and Photographs Division [LC-USZ62-95310])

the German mathematician who is best known as one of the two inventors of calculus. Leibniz had a deep interest in all matters philosophical, and this particular innovation was motivated by his interest in religion. He wanted a system of numeration that reflected his religious beliefs. He proposed a system of positional numeration that consisted of two symbols, 0 and 1. The symbols 0 and 1 would represent not just numbers but philosophical ideas. The 0 would represent the void; the 1 would represent God.

To illustrate how to count with just the two symbols we begin by (again) considering the system for counting soldiers that was first described in chapter 1. When the first soldier passes the counter a single pebble is placed in the 1s pile. When the second soldier passes the counter the 1s pile is wiped clean and a single pebble is placed in the 2s pile. As the third soldier passes the counter, a pebble is placed in the 1s pile. When the fourth soldier passes the counter the first two piles are wiped clean and a single pebble is placed in the 4s pile. The pattern can continue indefinitely. Every number, whether or not it is a whole number, can be expressed in base 2. From a computational viewpoint the advantage of base 2 is its simplicity. Its disadvantage is that even relatively small numbers can become unwieldy. The number that in base 10 is written as 100, for example, when expressed in base 2, has the form 1100100.

Leibniz's philosophical arguments for the adoption of base 2 numeration made little impression on his contemporaries. His idea gathered dust for more than two centuries, until the advent of digital computers. Computers are controlled by large collections of small circuits, each of which can be in one of two states, a *zero state* or a *one state*. Because most computations are now performed by computing machines rather than humans, we can see that after a slow start, Leibniz's idea about binary, or base 2, notation has triumphed.

With a true positional system of base 10 notation in hand, scientists and mathematicians were equipped with a tool of extraordinary power. Many computations that had previously required a great deal of time and effort were now almost trivial, and many computations that had once been too difficult to complete were now within reach. This development further accelerated progress in mathematics and science.

4

ANALYTICAL ENGINES

We have already remarked that arithmetic performed with positional numeration and decimal fractions is so simple that junior and senior high school students routinely perform the same kinds of computations that previously could be done only by experts. Actually our current system of decimal notation is even simpler than that. It is simple enough for a machine to use. Arithmetic operations in our current system of positional numeration are not



Before the advent of the computer, large-scale computations had to be broken into small, human-scale subproblems. (Library of Congress, Prints and Photographs Division)

just mechanical: They are mechanically simple. This is one of the great advantages of our system of numeration. In fact around 1619, at about the same time that John Napier was preparing his book *Mirifici Logarithmorum Canonis Constructio* for publication, a professor of astronomy, mathematics, and Hebrew at the University of Tübingen was designing a mechanical calculator. His name was Wilhelm Schickard (1592–1635).

Schickard's calculator could add and subtract automatically and, with some additional help from the user, could multiply and divide. Schickard was hoping to send his invention to Johannes Kepler, the author of Kepler's laws of planetary motion. Kepler's research required him to perform many calculations. Kepler and Schickard corresponded about the possibilities of such a device, but Schickard never completed his machine. The calculator was only partially completed when it was destroyed in a fire. Before Schickard could rebuild the device he died in an outbreak of bubonic plague. Because of these tragedies his work on the calculating machine had little influence on those who followed. His notes, which contain a detailed description of the machine, were rediscovered in the 20th century. His 20th-century successors at Tübingen University used the notes to construct a model of his work, which is on display at the Deutches Museum in Munich, Germany.

Shortly after Wilhelm Schickard invented his mechanical calculator another mathematician, Blaise Pascal (1623–62), built his own mechanical calculator. Pascal was unaware of Schickard's work. Instead Pascal's original intention was to help his father with the many calculations he performed in his work as a government official. Eventually Pascal made a number of these machines and sold some of them. A few still exist. The machines, called Pascalines, were designed for addition and subtraction; one could not use them for division and multiplication. By all accounts Pascal's machines were of a less sophisticated design than Schickard's. Their importance and their beauty lie in the concept behind the calculations. Mechanically they consist of a series of wheels with the numerals 0 through 9 on them. The wheels are connected to each other in such a way that when one wheel is rotated 360° the wheel immediately to the left rotates $1/10$ of a

CALCULATORS, COMPUTERS, AND THE HUMAN IMAGINATION

Schickard's, Pascal's and Leibniz's mechanical calculators share an important characteristic with even the most modern computers: The set of numbers that machines are designed to manipulate is fundamentally different from the set of numbers that we humans imagine.

Our number system, as we envision it, is infinite. To any number, however large, we can, in theory, always add a 1 and obtain an even larger number. And any positive number, however small, can always be multiplied by 1/2 to obtain a number that is still bigger than 0 but only half the size of the original. We can also add very large numbers and very small numbers together, and our answer will be different from both of the two numbers with which we begin. No person questions the truth of these obvious-looking statements, but no machine can implement them, either. No matter what number we enter into the Pascaline, if we repeatedly add a 1 to it, we eventually exhaust the capacity of the machine to add. This is also true of all the successors of the Pascaline as well.

This limitation did not make these mechanical calculators useless. As long as their use was restricted to the small collection of arithmetic problems that they were able to solve, the machines relieved the user of much of the tedium associated with routine calculation. This is precisely what they were designed to do. Still the limitations of the early calculators must have been evident to even the occasional user. Mechanically Pascal, Schickard, Leibiz, and others had found a way to represent part—but only part—of our number system through a system of gears and levers. There would, however, never be enough gears and levers to represent all of the numbers that they were able to imagine.

As mechanical calculators made way for much more powerful electronic computers, gears and levers gave way to transistors. But the limitations of these new machines, though less obvious to most users, are still present. The limitations still arise from the very fundamental difference between our infinite system of numbers and its finite representation inside the machine. In one form or another this difference between machines and the human imagination is maintained to the present day.

revolution. This is the mechanical version of “carrying the 1.” Here we see an example of our positional system of numeration represented not with 10 symbols but with wheels and gears. This

is the advantage of our system of numeration. It is so mechanical that it can be represented by gears and levers or even electronic circuits!

The Pascaline could also add decimal fractions as well as whole numbers. Because we carry the 1 when adding decimal fractions in just the same way we carry the 1 when we add whole numbers, the Pascaline is just as capable of adding decimal fractions as it is whole numbers. The key is to enter each number in the Pascaline with the same number of digits to the right of the decimal point. For example, if one wanted to add the numbers 3.41 and 11.1, one would need to enter 3.41 and 11.10 into the Pascaline and then remember to interpret the answer as 14.51 instead of 1451. This is entirely equivalent to “lining up” the decimal points when we add a list of numbers by hand.



This mechanical calculator was constructed from the notes of the University of Tübingen professor Wilhelm Schickard (1592–1635). (Courtesy of IBM Corporate Archives)

Pascal did not sell many of these calculators. There was not much demand. The manufacturing and assembly processes were crude, so the parts did not always fit together well. As a consequence they were not mechanically reliable. In addition to the mechanical problems, the machines were expensive. Nevertheless Pascal's was one of the first more or less successful efforts at mechanizing arithmetic processes.

For the next two and one-half centuries mathematicians and inventors worked at designing better versions of the Pascaline. Essentially all of these devices were the base 10 system of positional numeration represented as a series of interlocking gears. Inventors hoped that these devices would free people from the tedium of calculation.

One of the earliest and most notable improvements on Pascal's design was made by Gottfried Wilhelm Leibniz. Leibniz was one of the most versatile and creative mathematicians in history. We noted in chapter 3 that Leibniz developed the binary system of numeration and coinvented calculus. Leibniz acknowledged that the part of his machine that performed additions was identical to the Pascaline, but his invention could also multiply. It did this through a device now known as a Leibniz wheel. Mechanically Leibniz's calculator represented multiplication as a series of additions. The process of multiplying one number by another was achieved by turning the handle of a crank. This invention influenced generations of designers.

Charles Babbage and the Analytical Engine

The next breakthrough in computation occurred when the British mathematician and inventor Charles Babbage (1791–1871) began to think about what he called an analytical engine. Charles Babbage's analytical engine was the forerunner of the modern computer. In some ways he anticipated the design of modern computers, although the language that he used to describe his ideas definitely is that of another age. For example, he imagined that numbers would be kept in the “store,” which,

today, we call the memory, and that the processing would occur in a different part of the computer, which he called the “mill.” What Babbage called the “mill” we call the central processing unit (CPU). He imagined that the computer would receive its instructions via punched cards, and in all of these ideas he generally anticipated how modern computers would be designed. (Of course modern computers do not use punch cards, but the first few generations of electronic digital computers could not work without them.) On the other hand he wanted to run his computer with a steam engine, which was the most modern power source of the era. (Steam-powered computers never caught on.) More importantly for us he thought that each number should be entered into the store and the mill with the same number of decimal digits.

It is important to understand Babbage’s ideas about how numbers should be represented in computers, because these ideas guided the first modern attempts at building electronic computers. Babbage envisioned that his engine would store and manipulate numbers expressed in our system of base 10, positional numeration. Each digit would be stored in a place, or “cage,” in the machine—one digit per cage. Because the mechanics of manipulating decimal fractions are exactly the same as those involved in manipulating whole numbers, Babbage said that if one wanted to work with decimals one could just imagine where the decimal point would be and then interpret the results accordingly. We call this method of representing numbers fixed-point notation



A precursor to the never-finished analytical engine, Charles Babbage's difference engine was designed to perform computations associated with astronomy, navigation, and engineering applications. (Courtesy of IBM Corporate Archives)

because the decimal point remains in the same place throughout the computation. In other words if one number in the machine contained a 10ths column then *every number* entered into the machine should contain a 10ths column, even if the 10ths entry were a 0. Babbage did this so that all decimal points would be lined up in much the same way that we line up the decimal points whenever we do arithmetic by hand. Babbage's analytical engine, as the Pascaline, could only operate with the decimal points aligned. If every number were entered into the analytical engine with the same number of decimal digits, then the user was only required to remember where the decimal point was located initially, because the result was also expressed with the same number of decimal digits. The engine kept track of the rest of the arithmetic.

One difficulty with fixed-point arithmetic is that it severely limits the quantity of numbers with which the machine can work. For example, suppose we decide to store each number in four of what Babbage called cages. (Babbage proposed a larger number of

AN EARLY ELECTRONIC REPRESENTATION OF OUR NUMBER SYSTEM

Early research papers on computer design show that although researchers were making use of new electronic technologies, their conception of numbers was still old-fashioned. Initially they organized the circuits so that the computer could store numbers and perform arithmetic operations in our base 10 system of positional numeration. They still favored Babbage's fixed-point arithmetic as well. Consider this excerpt from an early paper that explains one way of representing numbers and performing additions and subtractions:

There are assemblages of (*ten*) thermionic tubes with associated circuits, having the following properties. One tube of the ring alone is in conducting condition. On receipt of a pulse this tube becomes unexcited, and the next in order becomes excited. The direction of succession around the ring, positive or negative, is determined by which one of a pair of associated

cages, but the principle is exactly the same.) Each of the cages holds exactly one digit, so an analytical engine with four cages can manipulate only three-digit numbers. The difference between the number of cages and the number of digits arises because the first cage holds the sign of the number: The first cage indicates whether the number is positive or negative. The next three cages must hold the digits that most closely approximate the number that we have in mind. If we deal in whole numbers this means that we can write any whole number from -999 up to +999. If, however, we decide to use two decimal places for increased accuracy, then, *because every number must also be represented with those same two decimal places*, we can store only numbers from -9.99 to +9.99. The range of numbers that we can represent is 100 times smaller! Part of the difficulty arises from the fact that we are trying to represent our infinite number system in a computer with finite storage capacity, but another source of difficulty lies in the fixed-point system itself. With fixed-point arithmetic what we gain in

electronic switches is energized. The ring thus counts pulses. A restoring circuit is provided, energizing of which restores the ring to the condition where the zero tube only is energized. . . . An activating circuit may also be associated with such a ring. When this is energized the ring steps automatically at a predetermined rate and in a predetermined direction back to the zero position, and on each step emits a pulse into a receiving circuit.

(*Bush, Vannevar. "Arithmetical Machine." Vannevar Bush Papers. Container 18, Folder: Caldwell, Samuel, 1939–1940. Library of Congress, Washington, D.C.*)

This may sound familiar. You have encountered this idea earlier in our story. It is a vacuum tube version of the Pascaline! The first paragraph describes the electronic version of turning a ring on the Pascaline: Each digit is represented by one tube in the ring and the "excited" tube represents the stored digit. The second paragraph describes how to "borrow" and how to "carry the 1." The "emitted pulse" is the electronic equivalent of turning the gear that connects one wheel on the Pascaline to an adjacent wheel. Nearly 300 years later computer design was still heavily influenced by Pascal's simple mechanical calculator.

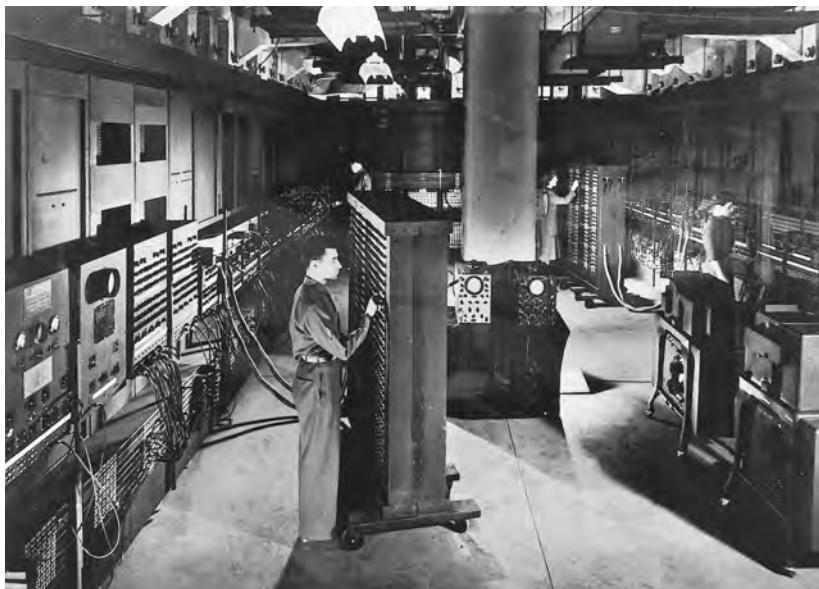
accuracy we lose in the range of numbers available for computation. The solution is not immediately obvious. Mathematicians and engineers devoted a lot of time to finding the most efficient way of representing our number system in a computer before arriving at the system presently in wide use.

Babbage never completed his machine. It seems that whenever he worked on it, he found a better way to design it and so would begin again from the beginning. He continued to refine his theory long after others lost interest in his project. Babbage was far ahead of his time in imagining how a computer should work. A lot of the architecture of a modern computer that we take for granted was first imagined by Babbage, but he was wrong about how numbers should be stored. Babbage's system of fixed-point arithmetic would lead to substantial difficulties in even relatively simple calculations. The problem is that Babbage's system is incapable of representing the wide variety of numbers that arise in many applications.

The Representation of Numbers in Computers

Computer design did not attract much attention until reasonably reliable vacuum tubes were developed in the first decades of the 20th century. The vacuum tube was the predecessor to today's transistor. Both tubes and transistors are designed to control and modify the flow of electrical current. Vacuum tubes, however, were slower to respond to changes in the circuit. They also required considerably more energy to perform the same function, but at the time vacuum tube technology seemed extraordinarily fast and efficient. Much of the original motivation for the development of computers stemmed from their ability to speed progress in computationally intensive fields, especially ballistics calculations, and later radar, sonar, and nuclear weapons research. One of the barriers to progress in all of these fields was the slowness with which calculations could be completed.

Computer design was greatly accelerated by World War II (1939–45). The first working electronic computer was called the Electronic Numerical Integrator and Computer (ENIAC). It was finished in 1946, one year after the end of World War II. ENIAC was a huge machine that contained 18,000 vacuum tubes. It was a



The Electronic Numerical Integrator and Computer (ENIAC) was the world's first digital computer. (Courtesy of John W. Mauchly Papers, Rare Book & Manuscript Library, University of Pennsylvania)

big step forward, but it continued to store and manipulate numbers in base 10 in a way that was somewhat similar to the way that Babbage proposed.

ENIAC and each of its immediate successors were individually designed and constructed. Each machine was very expensive and labor-intensive. Each represented a tremendous effort not just to build a computer but to overcome the limitations discovered in the previous machines. Almost as soon as ENIAC was built John von Neumann (1903–57), a prolific mathematician who was very involved in the theory and design of these machines, saw that base 2 represented a considerable gain in simplicity over the base 10 system that ENIAC had used. Because the circuits could assume only one of two states—the circuit was either off or on—base 2, the discovery of Gottfried Wilhelm Leibniz centuries earlier, was a more natural choice than base 10 for the computer. It simplified the design of the circuits.

Changing the base in which numbers are stored does not, however, address the problems associated with fixed-point arithmetic, and there was some controversy about how this problem could be solved. There were two choices: Designers could continue to work with fixed-point arithmetic, or they could incorporate what is called floating-point representation into the design of the machine. Floating-point arithmetic is more flexible in that it can better store numbers of widely varying sizes. There were those who believed that fixed-point arithmetic was the more efficient of the two methods of representing numbers. But other devices already being constructed in other places used the floating-point form, and the success of the new idea quickly convinced everyone that what we now call floating-point representation was the better of the two alternatives.

Floating-Point Representation

One of the fundamental jobs a computer has is to store numbers. The number may be obtained from an input file, or it may be the result of a computation that the computer has just finished performing. Very often the computer does not store an exact copy of this number in its memory. Instead it stores an approximation. This process is unrelated to the base in which the number is expressed. It relates to the limitations of the machine itself. Because a computer has only a finite amount of memory and there are infinitely many numbers, designers must be careful to ensure that the computer does not waste all of its memory storing a few or even just one very long number. We are usually unaware of how computers represent numbers because modern computers are so good at doing so that we do not usually notice their shortcomings. Of course many numbers can be stored in the computer's memory with perfect precision, but most cannot. The problem that concerns us arises when the computer stores a number that requires many digits to be expressed. For example $1/3$ as a decimal fraction is $0.333\dots$. The sequence of 3s does not terminate. Another example is a very large integer that has many nonzero digits. It may be a number so large that there would not be room enough in all of

the books in all of the libraries of the world to write. These kinds of numbers represent an additional challenge to computer designers because they *cannot* be stored with perfect accuracy. There will never be enough memory to do so.

The difference between the number that we want to store and the number that the computer actually places in memory is called the round-off error. Round-off error is unavoidable because the computer designer can allocate only a fixed amount of storage space for each number. Although there is great variety in the numbers to be stored, the amount of storage space is the same for every number. Insofar as is possible, this storage space must hold all the vital facts about the number it contains. These facts include

- The sign of the number (i.e., whether it is positive or negative)
- The digits that make up the number
- The size of the number (Remember that in a positional system of numeration it is the position of each digit that tells us its value. Representing the digits that constitute the number and the size of the number are distinct problems.)

When accomplishing these tasks, computers must also represent the number in a form convenient for storage. We can think of this stored form of the number as a kind of scientific notation. When we write a number in scientific notation we write it as a product of two numbers. The first number is greater than or equal to 1 and less than 10. The number, called the mantissa, gives us information about the digits that make up the original number. The second number is a power of 10. The power of 10 that appears gives us information about the size of the number. Consider, for example, the number 172, which, when written in scientific notation, is 1.72×10^2 . The number 1.72 is called the mantissa. The superscript 2 is called the exponent. One way of understanding the exponent is that it is an indicator: The exponent indicates how many places we must move the decimal point in the mantissa to

FLOATING-POINT ARITHMETIC AND YOUR CALCULATOR

$$100,000 + 0.00001 - 100,000 = x$$

$$100,000 - 100,000 + 0.00001 = x$$

Although these two equations have the same solution most calculators produce distinct solutions because of the way that numbers are represented in machines.

Here is a simple example of how computers solve certain problems in a way that is fundamentally different from ours. Consider the following two equations.

The author solved both equations with his “scientific” programmable calculator. These two equations differ only in the order in which the operations are performed. A quick check should convince the reader that the order of

operations should not matter in this case. Both equations have the same answer. The answer should be 0.00001 for both equations, but in the first equation the calculator yielded the wrong answer of 0, and in the second equation the calculator gave the correct answer of 0.0001. The programmable calculator that was used works on the same principles as a more powerful and more sophisticated computer.

regain our original number. Alternatively we can think of the exponent as a measure of the size of the number: When the exponent is positive we move the decimal point to the right; when the exponent is negative we move the decimal point to the left. (Computers, as already noted, do not use base 10 to store numbers. We use base 10 here because most readers are more familiar with base 10 and because the principle is the same for any base.)

We can get some feeling for the way this works in practice if we imagine the storage space allocated for a number to be a series of boxes, or, as Babbage would say, cages. In each box we place one fact about the number. Part of the storage space contains information about the sign of the number—specifically whether the number is positive or negative—and part of the storage space contains information about the digits that make up the mantissa. Finally, part of

Here is why the programmable calculator gave two different answers to what was essentially the same problem. The numbers were entered into the calculator from left to right. In the first case the computer added 100,000 and 0.00001 to obtain 100,000.00001. Notice how long the mantissa is in base 10: It is 11 digits long. When the calculator stored this number in memory, the actual mantissa, which in base 10 we write as 1.0000000001, was too long to fit into the space allocated for its storage. The calculator responded as it was programmed to do: It rounded off the mantissa to 1. So after the mantissa is rounded off, the result of the first operation is lost. It is as if it never occurred. As far as the calculator is concerned $100,000 + 0.00001$ is exactly the same as 100,000. Next the calculator subtracts 100,000 from 100,000. The result is 0.

In the second equation we begin by subtracting 100,000 from 100,000. The result, which is stored in memory, is 0. Now we add 0.00001 to obtain the correct answer of 0.00001. To us the two equations are equivalent. To the author's calculator they are not! It is not hard to make up problems for which a more powerful computer will likewise yield the wrong answer. This is a fundamental problem that stems from the difference between our infinite number system and the finite number system used by all computers. The number system that we imagine is not the same system that is implemented on our machines.

the storage space contains two pieces of information about the exponent: the sign of the exponent and the value of the exponent.

To see how this works, imagine storing a number in a computer. We can imagine that our “number storage compartment” is designed to hold 10 facts about our number. All the information that the machine retains about our number must fit inside those 10 places. There are no exceptions. If there is not enough room to fit all of the necessary facts then we have to omit enough facts to obtain a fit.

First, imagine an empty number storage compartment:

||||| | | | | |

Notice that there are 10 compartments, or cages, to hold all the facts about our number. For purposes of illustration we store the

WHY COMPUTERS?

This chapter has been principally concerned with important technical questions about how computers store and work with numbers. It has not addressed the larger question of why computers were developed. Computers can now be found in libraries, schools, homes, businesses, and virtually all government offices and laboratories. Less obvious are the computers that are present in automobiles, planes, trains, and phones. Many of us use computers for research, recreation, and communication. None of these uses, however, reflects the original motivation for developing these machines.

Early computers were completely unlike the machines with which we are familiar. They were highly specialized. They were difficult to build and maintain. They were extremely expensive. Programming them was strictly for experts; high-level programming languages had yet to be invented, and so the art of computer programming consisted of writing very long, detailed sets of instructions expressed in sequences of 0s and 1s. The instructions were communicated to the machine via stacks of punched cards. Even the experts frequently programmed them incorrectly: A single misplaced 0 or 1 in a very long instruction sequence would often cause the program to halt until the incorrect digit was identified and corrected. Furthermore computers were stand-alone machines. They did not “communicate” with one another. The earliest version of the Internet, an innovation of the United States Defense Department, would not be invented until several decades after the first computers had been constructed. To appreciate who built computers and why, here is a list of some of the most famous and most successful first-generation machines:

1. **The Colossus.** This series of special-purpose computers—the same name was given to several different versions of the machine—was an innovation of the British government. Created during World War II at a secret government institution at Bletchley, England, the machine was designed to assist in the decryption of German messages. The Germans had developed an extremely sophisticated device, called the Enigma machine, to encrypt their messages. They thought that their codes were unbreakable, but the British, with the help of the Colossus and some of the best mathematical minds of that generation, made substantial headway in decrypting German communications.

2. The Electronic Numerical Integrator and Calculator (ENIAC) was also developed during World War II. Developed at the University of Pennsylvania and completed in 1946, it was originally called Project PX, and, as the Colossus was, was originally kept secret. The idea and much of the design work for this general-purpose digital computer were those of John W. Mauchly of the University of Pennsylvania, but technical expertise was drawn from several academic and government institutions. The money for ENIAC came from the United States Army. The effort was motivated by problems arising in the field of ballistics, the science concerned with the study of the propulsion and motion of projectiles. When propellant is ignited in a cannon, for example, huge forces are generated; shock waves are created; the projectile exiting the cannon is buffeted behind by gas rushing out of the cannon and buffeted ahead by the drag of the atmosphere. Predicting these factors and their effect on the projectile's path is an extremely difficult mathematical problem. Because of the demands of the war, the army was struggling with a backlog of ballistics computations when the proposal for a digital computer was called to their attention. They recognized how important such a machine could be—provided it could be made to work. They contributed technical help and \$500,000 to the effort.
3. The Electronic Discrete Variable Automatic Computer (EDVAC), the successor to ENIAC, was completed in 1950. In order to meet the urgent need for better wartime calculations, the design characteristics of the ENIAC were fixed early in the project so that work could proceed as fast as possible. As engineers and mathematicians worked on ENIAC they soon realized that there were better ways of designing a computer. Almost as soon as the ENIAC was completed, the Ordnance Department—now known as the Ordnance Corps—of the United States Army signed a contract with the University of Pennsylvania to create a second computer, the EDVAC. The army's motivation—and they contributed almost \$500,000 to this project as well—was the same as for the ENIAC: better and faster ballistics calculations.

(continues)

WHY COMPUTERS?

(continued)

4. The Ordnance Variable Automatic Computer (ORDVAC) was another first-generation computer. Developed by the United States Army and the University of Illinois and completed in 1951, it, too, was created for the purpose of doing ballistics computations.
5. The Universal Automatic Computer (UNIVAC) was created in 1951. Mauchly, one of the principal designers of ENIAC, was also one of the principal designers of UNIVAC. As the other computers on this list were, the UNIVAC was created for government use, but unlike the ENIAC, EDVAC, and ORDVAC, UNIVAC was made for the United States Census Bureau. It was the first major computer project created for peaceful purposes.

As this brief list makes apparent, computers were originally created by governments as part of the war effort. The U.S. Census, an organization that has always been on the forefront of technological change, was the one major exception. Relatively easy-to-use computers—the sources of recreation and information with which we are familiar—would not be available until a generation after the United States Army began to commission machines for help in ballistics calculations.

number 123,456,789. (In scientific notation we write 123,456,789 as 1.23456789×10^8 .)

The first fact we store is the sign of the number. The next six places hold facts about the mantissa. The place immediately to the right of the mantissa holds the sign of the exponent, and finally we keep two places for information about the exponent. We store the number in four steps:

- We begin by placing a + sign in the first position because our number is positive. Now our compartment looks like this: |+| | | | | | | | | |.

- Now we place *as much of the mantissa as we can fit* into the next six slots. At this point our compartment looks like this: `|+|1|2|3|4|5|7| | | |`. (Notice that there is not room enough to fit the entire mantissa, so we round it off and use the first six digits. Notice, too, that because every mantissa is greater than or equal to 1 and less than 10, it is not necessary to store the decimal point because it is always in the same place—immediately after the first digit of the mantissa.)
- The exponent is a *positive* 8, so we place a + sign in the next box: `|+|1|2|3|4|5|7|+| | |`.
- Finally, because the exponent is 8 we write a 08 in the remaining two empty boxes. Our number compartment looks like this: `|+|1|2|3|4|5|7|+|0|8|`. (Notice that other numbers, for example, the numbers 123,457,000 and 123,456,955, are also rounded off to exactly the same number as that stored in our imaginary computer.)

One great advantage of floating-point notation is that it is adaptive. It can store numbers of widely varying sizes. It is also more useful for storing *significant* figures, and the more positions set aside to store the mantissa the less likely it is that important information will be lost. In fact the precision of the machine is often measured by the number of “cages” set aside to store information about the mantissa.

Floating-point notation has proved to be a much more flexible system than the fixed-point notation that was envisioned by Babbage and implemented in the Electronic Numerical Integrator and Calculator (ENIAC). Floating-point notation is not perfect (see the accompanying sidebar), but it is so far the best method available for representing our infinite set of numbers in a machine that can store only a finite amount of information. After its introduction in the 1950s it quickly displaced most fixed-point applications. Today desktop computers and supercomputers alike use a floating-point system to store and manipulate numbers.

Although most people continue to use the strictly positional, base 10 system of notation employed by Napier about four centuries earlier, this is not the system of notation most in use today. Because machines now perform the vast majority of all computations, floating-point notation—a system with which very few of us have any direct experience—is probably more characteristic of our time than the decimal system of notation pioneered by Napier, Stevin, and Viète.

PART TWO

EXTENDING THE IDEA OF A NUMBER

5

AN EVOLVING CONCEPT OF A NUMBER

Knowledge of positive fractions and positive whole numbers, what we now call the positive rational numbers, is at least as old as civilization itself. The ancient Egyptians, Mesopotamians, and Chinese all found ways to compute with positive rational numbers. This is not to say that mathematically advanced ancient cultures had number *concepts* similar to ours. Although ancient computational techniques were sometimes identical to those taught in school today, ancient number concepts were often quite different from those with which we are familiar. Nor was the transition from ancient to modern concepts easy or swift. It is an extraordinary fact that the development of a clear understanding of negative numbers took longer than the invention of calculus.

Emotional as well as intellectual considerations proved to be significant barriers to progress. Numbers are fundamental—they are the building blocks and the principal source of examples that mathematicians use as they explore mathematics—and each generation of mathematicians, ancient and modern, is heavily invested in what it assumes numbers represent. Seeing beyond old ideas and preconceptions has always been difficult and time-consuming. This is as true in mathematics as in any other field.

Today we interpret numbers in the language of geometry. At an early age we learn about the “real number line.” The idea is simple enough: Imagine a line with two special points. The first point is 0. The 0 point divides the line into positive and negative parts. The second point is the number 1. The distance from 0 to 1

establishes a scale. The point that we associate with the number 2, for example, is located on the same side of 0 as the number 1 and exactly twice as far from 0 as the 1 point. The position of the number 1 on the line also determines a direction. Numbers on the same side of 0 as the number 1 are positive. Those located on the opposite side of 0 from the number 1 are negative. The position of every other number on the line is now determined. In this representation of the number system, every number represents a *directed distance* from 0; that is just another way of saying that the positive numbers lie to the right of 0, the negative numbers lie to the left of 0, and each number communicates information about how far the corresponding point is located from the origin. This simple and elegant representation of our real number system is a powerful conceptual aid. It is also a relatively recent historical innovation.

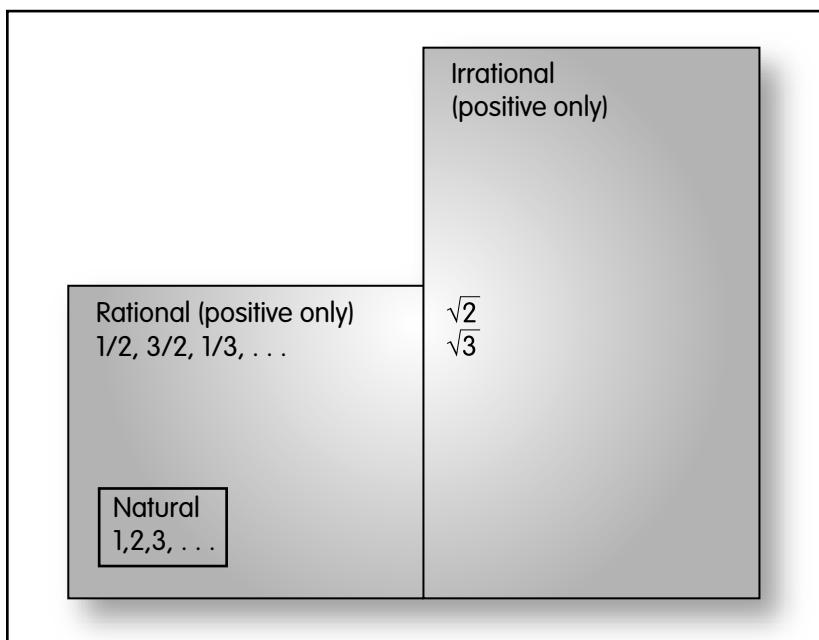


Diagram of an early concept of the number system. Mathematicians began to realize that without a broader definition of number, most algebraic equations could not be solved.

Earlier concepts of number did not depend on the real line. For the most part the numbers that early mathematically advanced cultures used were all positive. Geometrically they used only the numbers that are located to the right of 0 on the real number line. Most cultures simply refused to consider negative numbers or the number 0. Lacking a real number line, they identified numbers with objects or parts of objects. Sometimes the set of positive whole numbers sufficed: 10 people, 5 lakes, 3 mountains, and so on. Other times fractions were necessary: $3 \frac{1}{2}$ feet, $\frac{1}{4}$ kilogram, or $\frac{1}{2}$ day, for example. For most of human history no one took time to consider a set of no apples, or a collection of -3 days. Fortunately the collection of positive rational numbers was adequate for *most* purposes. We emphasize the word *most*, because there have always been indications that other numbers exist. There were even indications that they might be useful. Mathematically stumbling on other classes of numbers was almost unavoidable. To understand how classes of numbers outside the set of positive rational numbers were discovered, considering the operations that make up basic arithmetic is helpful.

There are four basic operations of arithmetic: addition, subtraction, multiplication, and division. In what follows we need to be concerned with a fifth operation as well, the extraction of roots. (By *extraction of roots* we mean finding square roots, cube roots, and so on.) Three of these operations presented no conceptual problems to early mathematicians. For example, when we add two positive rational numbers the result is another positive rational number. Similarly if we multiply or divide two positive rational numbers the result is still another positive rational number. Today mathematicians describe this situation by saying that the positive rational numbers are closed under addition, multiplication, and division. The word *closed* means that we cannot obtain a number outside the set by adding, multiplying, or dividing numbers chosen from inside the set. The situation is different with subtraction and the extraction of roots.

When we subtract 1 from 2 we obtain a positive rational number. The answer, of course, is the number 1. When we subtract 2 from 1, however, our answer is negative. Subtracting 2 from 1

leads us off the positive half of the real number line. Today we would accept the number -1 as an answer to the problem of subtracting 2 from 1, but earlier in history mathematicians often refused to consider the question. To them it seemed impossible to subtract 2 from 1. Problems that we answer with a negative number, they asserted were unsolvable. There was no mathematical reason for them to restrict their attention to positive numbers. This limitation was one that these mathematicians imposed on themselves. Nor was the only conceptual difficulty the existence of 0 and negative numbers.

In mathematics many more possibilities exist than simply whether or not a number is greater than 0 or less than or equal to 0. If, for example, we search for the square root of 2, written $\sqrt{2}$, we find a number that cannot be expressed as a positive rational number. The number $\sqrt{2}$ is positive, but it is not *rational*; that is just another way of saying that it cannot be expressed as a fraction with a whole number in the numerator and another whole number in the denominator. One more example: If we search for the $\sqrt{-2}$, we find a number that has no place anywhere on the real line. The $\sqrt{-2}$ exists, but not in the real number system. Earlier in history mathematicians would have classified the number we now write as $\sqrt{-2}$ as meaningless.

We now know that $\sqrt{-2}$ is not meaningless. It is an example of a complex number, and it has no position on the real number line. Complex numbers have important uses in science and mathematics. The search for the meaning of different types of numbers has occupied some of the best mathematical minds in history. The search for a number system that is closed under basic arithmetic and the extraction of roots was finally completed less than 200 years ago. Many of its basic properties were discovered even more recently.

Irrational Numbers

An *irrational number* is a number that cannot be written as a quotient of two whole numbers. Roughly 4,000 years ago the Mesopotamians first encountered the irrational number $\sqrt{2}$ when

they computed the length of the diagonal of a square with side of length 1. The Mesopotamians were extremely adept at calculation. They had the mathematical acumen to compute highly accurate *rational number* approximations to $\sqrt{2}$: That is, they could find rational numbers that are very close to the irrational number $\sqrt{2}$. Cuneiform tablets have been found with rational approximations to the $\sqrt{2}$ that are accurate to the 1,000,000th decimal place. This is an accurate enough estimate of the size of $\sqrt{2}$ for most applications—even today.

The algorithm used by the Mesopotamians was *recursive*: This means that it consisted of a series of steps that could be repeated as often as desired. The more often the steps were repeated the more accurate the estimate of $\sqrt{2}$ became. In particular if we specify that we want an approximation to be within some margin of error—suppose, for example, that we specify that we want an approximation for $\sqrt{2}$ that is within 1/1,000,000,000th of the true value—the Mesopotamian algorithm can find a rational number that lies within the specified tolerance. All that is necessary is to repeat the series of steps that make up the algorithm enough times.

This is an excellent algorithm for *approximating* $\sqrt{2}$, but it *cannot* be used to find an exact value for $\sqrt{2}$. The reason is that each step of the Mesopotamian algorithm consists of a sequence of additions, multiplications, and divisions. The initial input in the Mesopotamian algorithm was always a positive rational number and, as we have already pointed out, the positive rational numbers are closed under additions, multiplications, and divisions. As a consequence *every* sequence of additions, multiplications, and divisions of a set of positive rational numbers can produce only other positive rational numbers. Because the $\sqrt{2}$ is irrational there is always some distance between the $\sqrt{2}$ and the number computed with the algorithm.

To understand why the Mesopotamian algorithm can approximate irrational numbers with rational ones, some knowledge of the placement of the rational numbers is useful. Given any two numbers we can always find a rational number between them. It does not matter how close together the numbers are. As long as

the two numbers are distinct, there is a rational number that is greater than the smaller one and less than the larger one. Today mathematicians describe this situation by saying that the rational numbers are dense on the real number line: That is, there is no interval, however small, that does not contain a rational number.

The denseness of the rational numbers on the real line helps to explain why the Mesopotamian algorithm for computing $\sqrt{2}$ works so well. We can imagine surrounding the $\sqrt{2}$ with a small interval. Geometrically the “smallness” of the interval represents our margin of error. What the Mesopotamian algorithm allows us to do is to calculate one of the rational numbers that lie inside this interval, and there is always a rational number inside the interval because the rational numbers are dense on the real line. The Mesopotamians would not have understood their algorithm in these terms, because they would not have known how the rational numbers are placed on the real number line. Nor is there any indication that they were aware of the existence of irrational numbers. Nevertheless they made use of these facts whenever they employed their algorithm for computing square roots.

Although the Mesopotamian algorithm yielded only an approximation to $\sqrt{2}$, there is no indication that this “shortcoming” was a source of concern to them. Nor is there any indication that they were curious about why each step of their algorithm would yield a better approximation to $\sqrt{2}$ instead of $\sqrt{2}$ itself. Rather than searching for a perfectly exact answer, they simply stopped when their approximation was “accurate enough.” The degree of accuracy they required depended on the application that they had in mind.

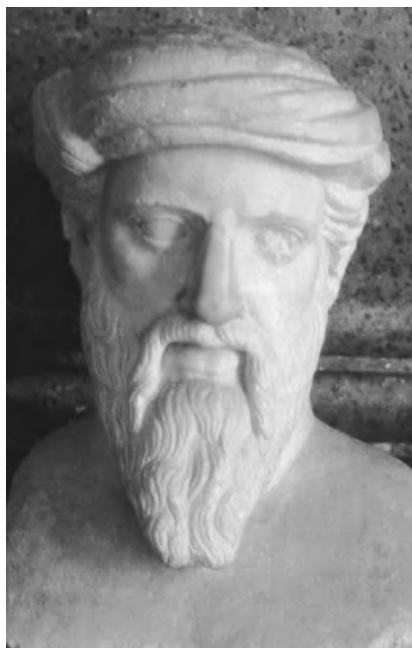
It may seem strange that they did not investigate the problem of obtaining an exact answer versus an inexact estimate. But if the idea of finding *only* an approximation for $\sqrt{2}$ seems not quite satisfactory, we should remember that our calculators are also capable of finding only rational approximations of irrational square roots. This is a consequence of the way calculators and computers store numbers. In fact the algorithm used by our calculators to compute square roots is often the very same algorithm that was used by the Mesopotamians!

Pythagoras of Samos

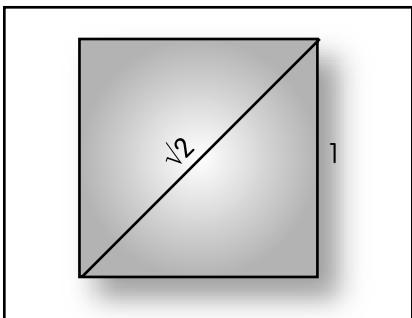
As math students we usually encounter positive irrational numbers for the first time when we compute square roots. Historically this was the way they were first uncovered as well. The credit for the discovery of irrational numbers is usually given to the Greek philosopher and mathematician Pythagoras of Samos (ca. 582–ca. 500 B.C.E.) and his followers. The Pythagoreans discovered that the number $\sqrt{2}$ is irrational. They learned something new about the nature of this number. There is no indication that they discovered anything new about the size of the number. The Mesopotamian estimate of the size of $\sqrt{2}$ was almost certainly more accurate than any estimate that Pythagoras ever knew.

In contrast to the work of the Mesopotamians, the discovery by the Pythagoreans that $\sqrt{2}$ cannot be represented as a quotient of two whole numbers reveals little about the size of the number. Still this is an important mathematical insight made by one of the most influential of all the Greek philosophers and mathematicians.

Pythagoras was a philosopher and a mystic with a penchant for travel. As a young man he visited Egypt and perhaps Mesopotamia, and he absorbed the philosophy and mathematics of these societies. He eventually settled in the Greek city of Cortona, located in what is now Italy. He was a charismatic person and was soon surrounded by followers. They built a settlement and lived apart from the rest of the community. Pythagoras and his followers



Pythagoras of Samos. He and/or his followers discovered the existence of irrational numbers. (ARPL/Topham/The Image Works)



Geometrically the irrationality of $\sqrt{2}$ means that there does not exist a line segment of length x , where x can represent any number, and whole numbers m and n , such that $mx = 1$ and $nx = \sqrt{2}$. In other words, the side of a square (of length 1) and its diagonal (of length $\sqrt{2}$) share no common measure.

numbers that are not rational: that is numbers that cannot be expressed as the quotient of two whole numbers. Their discovery disappointed and unsettled them.

Mathematics was a central part of Pythagorean philosophy. “All is number” was their maxim. But when they asserted that all is number they were referring only to a very specific kind of number. They believed that all relationships in nature could be expressed as ratios of whole numbers. When they discovered the existence of irrational numbers, numbers that *cannot* be expressed as ratios of whole numbers, they also discovered that their concept of number and its role in nature was flawed. Because these ideas were central to their philosophy, they tried—unsuccessfully—to conceal their discovery.

The discovery of the irrationality of $\sqrt{2}$ soon led to the discovery of a larger set of irrational numbers. The Greeks soon composed a short list of numbers that they had proved were irrational, but aside from discovering what irrational numbers are *not*, the Pythagoreans made relatively little headway into discovering what irrational numbers *are*. They knew that irrational numbers *cannot* be represented as a quotient of two integers, but identifying something by what it is not

lived communally, sharing everything, owning nothing individually. Their philosophy of communalism also applied to claiming credit for new ideas and new mathematical discoveries. As a consequence we cannot distinguish between what Pythagoras discovered as an individual and what his disciples learned on their own. Probably Pythagoras and his disciples would have rejected that kind of he—they distinction anyway. We can, however, be certain of this: It was the Pythagoreans who first learned of the existence of

that is numbers that cannot be expressed as the quotient of two whole numbers. Their discovery disappointed and unsettled them.

Mathematics was a central part of Pythagorean philosophy. “All is number” was their maxim. But when they asserted that all is number they were referring only to a very specific kind of number. They believed that all relationships in nature could be expressed as ratios of whole numbers. When they discovered the existence of irrational numbers, numbers that *cannot* be expressed as ratios of whole numbers, they also discovered that their concept of number and its role in nature was flawed. Because these ideas were central to their philosophy, they tried—unsuccessfully—to conceal their discovery.

The discovery of the irrationality of $\sqrt{2}$ soon led to the discovery of a larger set of irrational numbers. The Greeks soon composed a short list of numbers that they had proved were irrational, but aside from discovering what irrational numbers are *not*, the Pythagoreans made relatively little headway into discovering what irrational numbers *are*. They knew that irrational numbers *cannot* be represented as a quotient of two integers, but identifying something by what it is not

THE IRRATIONALITY OF $\sqrt{2}$

The proof that $\sqrt{2}$ is irrational is one of the most famous proofs in the history of mathematics. The Greek philosopher Plato (ca. 427 B.C.E.–347 B.C.E.) was so amazed by the existence of irrational numbers that he divided his life into that part that occurred before he saw the proof and that part that occurred after. Plato was not alone. The existence of irrational numbers surprised and fascinated the Greeks.

The proof is a nonexistence proof. To show that $\sqrt{2}$ is irrational we must demonstrate that there are no integers a and b such that $\sqrt{2}$ equals a/b . The proof depends upon two facts about even numbers:

- 1.** If a^2 is divisible by 2 then $a^2/2$ is even.
- 2.** If b^2 (or a^2) is divisible by 2 then b (or a) is even.

To show that $\sqrt{2}$ is irrational we begin by assuming the opposite: Suppose that $\sqrt{2}$ can be represented by a fraction a/b where a and b are positive whole numbers. If a and b exist then we can also assume that *the fraction a/b is in lowest terms*. This is critical, but it does not constitute an additional restriction. If, after all, a/b were not in lowest terms, then we could certainly reduce it until it was, so we might as well begin by assuming that a/b is already in lowest terms. Notice that when a/b is in lowest terms, it cannot be the case that both a and b are even numbers. (There is, however, nothing to prevent one of them from being even, but if both are even, then a/b is not in lowest terms.) Here, in modern notation, is how the ancient Greeks reasoned:

Suppose $a/b = \sqrt{2}$

Now solve for b to get

$$a/\sqrt{2} = b$$

Finally, square both sides

$$a^2/2 = b^2$$

This completes the proof. All that is required is to read off what the last equation tells us. The left side is even. (See the first fact.) This shows that b^2 is even. (If one side of the equation is even so is the other.) We conclude that both a and b are even. (See the second fact.) That *both* a and b are even numbers contradicts our assumption that a/b is in lowest terms. We have to conclude that our assumption that $a/b = \sqrt{2}$ is false. This proves that a and b do not exist.

(continues)

THE IRRATIONALITY OF $\sqrt{2}$

(continued)

Notice that this proof tells us virtually nothing about the number called $\sqrt{2}$. It is an assertion about what $\sqrt{2}$ is not: It is not rational. Nevertheless this proof is probably one of the most important in the history of mathematics. Irrational numbers continued to perplex mathematicians for the next few thousand years. Their existence was the first indication that numbers could be far more complicated and less intuitive than anyone had imagined.

does not yield much insight into the object itself. This type of insight could not have been very satisfying to the Pythagoreans, a group of individuals whose life was entirely devoted to the study of philosophy.

Although the discovery of irrational numbers was important from a philosophical point of view, from a practical point of view the discovery of the positive irrational numbers had little effect on computation. The reason, as was previously stated, is that the rational numbers are dense on the real number line. Any computation that involves irrational numbers can be performed with rational numbers, and the difference in the two solutions can be made as small as we please.

Yet the Pythagorean discovery that $\sqrt{2}$ is irrational had important theoretical implications about the existence of a larger, more complete number system. Unfortunately the Pythagoreans did not follow up on their discovery. They continued to work with positive rational numbers. In fact throughout the history of Greek mathematics most mathematicians refused to consider irrational solutions to problems whose solution depended on computation. Diophantus of Alexandria (ca. third century C.E.), the most successful of all Greek mathematicians who were interested in algebra, did not consider irrational solutions acceptable. He posed and solved many problems, but in each case he restricted his attention to solutions that consisted of rational numbers. The discovery of positive irrational numbers hinted at the existence of a larger, more complex number system, but for a very long time no further progress was made in understanding the nature of the number system.

6

NEGATIVE NUMBERS

Whenever a larger number is subtracted from a smaller one, a negative number is the result. Geometrically subtracting a large positive number from a small positive number moves us to the left of 0 on the real number line. Consequently it is hard to avoid occasionally stumbling upon a negative number in the course of solving a problem. This is especially true if the problem requires extensive calculations, because even if the answer is positive, some intermediate steps in the calculation may yield negative numbers. Nevertheless many early cultures did not use negative numbers.

One of the earliest of all cultures to make regular use of negative numbers was the Chinese. In the first section of this volume we discussed the Chinese method of numeration, which used rod numerals. Rod numerals looked as if they had been formed by straight rods, and sometimes that was, in fact, the case. Colored rods, called counting rods, were used as a computation device in ancient China. The user generally carried two sets of rods, one red and one black. The red rods represented positive numbers. The black set represented negative numbers. These were combined according to certain well-defined rules to obtain answers. Such a concrete and elegant system for performing calculations makes the existence of negative numbers easy to accept. Nevertheless the ancient Chinese were cautious. They used negative numbers as intermediate steps during the course of their calculations. They did not accept negative numbers as final solutions to problems. The first great, radical breakthrough in the concept of number occurred in ancient India.



For centuries the ancient astronomical observatory at Ujjain, India, was a center of mathematical research. It has many exotic-looking architectural elements that are used for astronomical observations.

(Dinodia/The Image Works)

As is the concept of 0, the concept of a number system that fully incorporates negative numbers is an idea we owe to Indian mathematicians. Unlike the concept of 0, which cannot be attributed to any individual mathematician or mathematical text, negative numbers can be attributed to the writings of the Indian mathematician and astronomer Brahmagupta (ca. 598–ca. 655), who wrote an early, complete mathematical treatment.

Not much is known of Brahmagupta's life. He was head of the astronomical observatory at Ujjain, which was, for several centuries, one of the great centers of astronomical and mathematical research in the world. His major work is *Brahma Sphuta Siddhānta*. It is principally an astronomy text and was written, in part, for the purpose of correcting and modernizing an earlier astronomy book, *Brahma Siddhānta*. In addition to information about astronomy Brahmagupta's book contains several chapters on mathematics. In those chapters Brahmagupta explains how to incorporate the number 0 and the negative numbers in arithmetic.

Brahma Sphuta Siddhānta contains little of the stark beauty of Greek texts. Numerous statements are made without proof or other supporting arguments. This may have been the result of the literary form Brahmagupta used to convey his ideas. As was the custom in his time and culture, the book was written entirely in verse. In any case and for whatever reasons, Brahmagupta shows us only his conclusions, which are written in the form of a series of rules. He provides no justification for these rules; nor does he reveal how the rules were discovered.

The concept of number described in *Brahma Sphuta Siddhānta* is a huge step forward. In the section “Algorithm” we find for the first time most of the additional concepts necessary to add, subtract, multiply, and divide positive and negative numbers as we do today. Essentially Brahmagupta expands the set of numbers so that no matter how we add, subtract, multiply, or divide these numbers, we always get an answer that is also in this new and expanded set. Mathematically speaking that is the innovation. Practically speaking his innovation is to expand the set of numbers to fit the problems he wants to solve, rather than to restrict the problems he is willing to consider to those with numerical answers he is willing to accept.

Was Brahmagupta aware of how different his mathematics was from the mathematics that had been developed in Mesopotamia, Egypt, China, and Greece? Perhaps. Hindu mathematicians of this era had some awareness of the mathematical accomplishments of the Greeks and possibly of the Mesopotamians and Chinese. Brahmagupta may have appreciated how these new ideas facilitated calculation and how their concept of number made available new solutions to old problems. He does not, however, include any comparisons with other cultures or provide the reader with any explanations that would help us understand how he and his contemporaries had expanded their concept of number.

Here are Brahmagupta’s own rules for subtraction. Note that “cipher” is what Brahmagupta calls 0:

The less is to be taken from the greater, positive from positive; negative from negative. When the greater, however, is subtract-

ed from the less, the difference is reversed. Negative, taken from cipher, becomes positive; and affirmative (*taken from cipher*), becomes negative. Negative, less cipher, is negative; positive (*less cipher*), is positive; cipher (*less cipher*), nought. When affirmative is to be subtracted from negative, and negative from affirmative, they must be thrown together.

(*Brahmagupta. Algebra with Arithmetic and Mensuration.*
Translated by Henry Colebrook. London: John Murray, 1819)

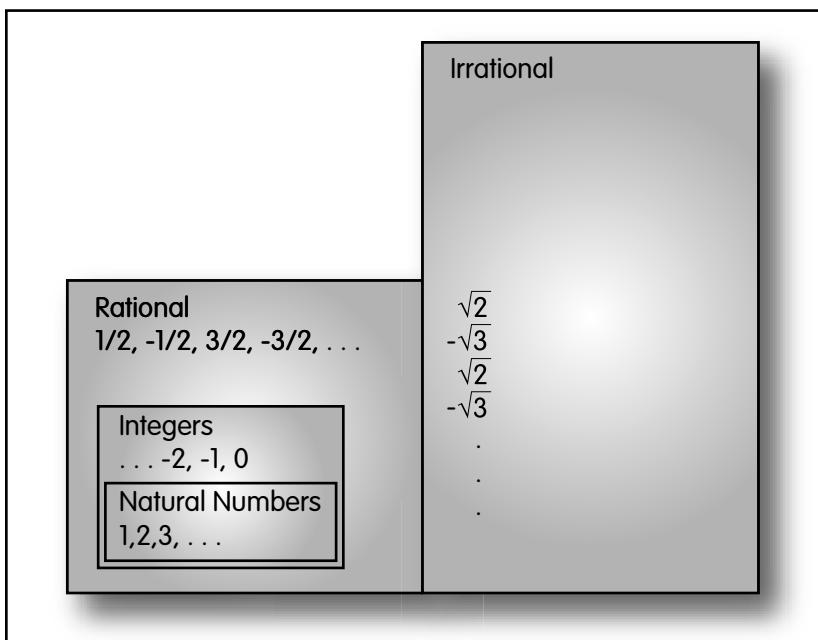
There are similar rules for the other arithmetic operations. In fact there are several more rules than those to which we are accustomed. (He includes rules for operating with irrational numbers, for example.) His list of rules is followed by a number of problems for the reader to solve.

Although he understands how to perform arithmetic with signed numbers, Brahmagupta encounters problems with the number 0. He clearly recognizes 0 as a number, and he recognizes it as a very special number; notice that in his rules for subtraction, quoted in the excerpt, 0, or “cipher,” is always treated separately. But Brahmagupta’s insight fails him when he attempts to devise a rule for dividing by 0. Here is his division rule:

Positive, divided by positive, or negative by negative, is affirmative. Cipher, divided by cipher, is nought. Positive, divided by negative is negative. Negative, divided by affirmative, is negative. Positive, or negative, divided by cipher, is a fraction with that for denominator: or cipher divided by negative or affirmative.

This rule correctly describes how signed numbers should be divided one into another, but he is wrong about 0. Today we know that division by 0 has no meaning. It cannot be defined without introducing logical contradictions. (See the section “Infinity as a Number” in chapter 10 to get some insight into the difficulties involved.) In particular his assertion that 0 divided by 0 equals 0 is false. Moreover he is more than a little vague about the

possibility of dividing a non-0 number by 0. To say that “*Positive, or negative, divided by cipher, is a fraction with that for denominator*” is to say almost nothing. In any case he does not reject the possibility of dividing by 0, which is the modern—and a logically consistent—solution to the problem. Conceptual difficulties involving arithmetic with 0 were not unique to Brahmagupta, however. Division by 0 continued to pose problems for mathematicians long after the publication of *Brahma Sphuta Siddhānta*. The English mathematician John Wallis (1616–1703), for example, one of the most accomplished mathematicians of his time, asserted that a positive number divided by 0 equals “infinity” and went on to conclude that if we divide that same positive number by -1 , it must be “larger than infinite.” Brahmagupta’s understanding of the arithmetic of signed numbers was not surpassed for about 1,000 years.



Islamic mathematicians and many European mathematicians of the Renaissance resisted expanding the number system to include the idea of negative numbers, but Hindu mathematicians used them freely.

ANCIENT MATHEMATICAL TEXTS FROM THE INDIAN SUBCONTINENT

The Hindu mathematical tradition occupies a special place in the history of mathematics. One of the earliest Hindu mathematicians whose work survived to modern times is Aryabhata I (476–550). His book is *Arya Bhateeyya* (also called *Aryabhatiya*). Because Aryabhata's work is fairly advanced, it apparently was preceded by many generations of mathematical research. Few records exist before Aryabhata's work, however, and so little is known about the history of mathematics on the Indian subcontinent before he wrote *Arya Bhateeyya*.

There is a long history of Indian mathematics after Aryabhata. It is often said that the Hindu mathematical tradition culminated with the work of the mathematician and astronomer Bhaskara II (1114–1185). During the six centuries that separate Aryabhata I from Bhaskara II there are a number of highly creative mathematicians and a number of important mathematical innovations. Despite the breadth and longevity of this tradition of mathematics, however, some characteristics remained the same from generation to generation and distinguish this tradition from other mathematical traditions.

First, in the Indian mathematical tradition mathematics and astronomy are very tightly intertwined. In fact Indian mathematicians produced very few texts on mathematics as a discipline separate from astronomy. Most mathematical writing appears as separate chapters in astronomy texts. Second, in this tradition almost all mathematics is expressed as a formal type of poetry, and that style may account for the terse form of presentation. Third, Hindu texts generally presented mathematical results without proofs. This characteristic is critical. A proof does not just assure the reader that a particular statement is correct. A proof is an aid to the reader who wants to understand the idea behind the statement better. The absence of proofs from most Indian works makes some of the statements found in these texts considerably harder to appreciate. Without a proof the reader who is unsure of what a particular verse means has little recourse other than to read the verse again. Without a proof the mathematician–author who is uncertain of the correctness of a particular statement lacks the ultimate mathematical tool for distinguishing the true from the false. Despite this shortcoming the ideas of the Hindu mathematicians were new and far-reaching, and their results had a profound effect on the development of mathematics around the world.

OUT OF INDIA

Some but not all Hindu ideas about number eventually found their way out of India. In the ninth century Baghdad was a center for learning and scholarship. There were Islamic, Jewish, and Christian scholars working at the great Islamic institutions of higher education. Mathematicians were, in addition to producing their own research, busy translating the mathematical texts of other cultures into Arabic. Translations of the major mathematical works of the Hindus and Greeks were much sought after. Arab rulers worked to obtain original texts, sometimes from other governments, which were then distributed to Arab scholars for translation and study. It was during this time in this environment that al-Khwārizmī, a mathematician and astronomer living in Baghdad, obtained a copy of a Hindu work. The exact work is no longer known, but some scholars speculate that it was Brahmagupta's *Brahma Sphuta Siddhānta*. Al-Khwārizmī does not indicate his source. In any case he used this text to write an exposition of the Hindu system of numeration. Unfortunately all copies of al-Khwārizmī's book in the original Arabic have now been lost. Only a single Latin translation of al-Khwārizmī's Arabic exposition of a Hindu text, *De Numero Indorum* (Concerning the Hindu art of reckoning), has survived into our own time.

It was al-Khwārizmī's book that introduced the Hindu system of numeration into Islamic culture, and it was through cultural contact with the Arabs that Europeans eventually learned of the Hindu system. Because the Europeans learned of the Hindu system from the Arabs, they attributed the system to the Arabs, although it is clear from the title of al-Khwārizmī's own book that the system is of Indian origin. If al-Khwārizmī did use Brahmagupta's book, he was very selective about what he chose to describe. Al-Khwārizmī was a prominent scholar, whose research into algebra influenced generations of Islamic and European mathematicians, but his conception of number was surprisingly limited given that he seems to have had access to Hindu mathematical texts. Although he recognized the importance of place-value notation, he did not share Brahmagupta's very broad understanding of what a number is. Al-Khwārizmī and other Islamic mathematicians preferred to work with positive numbers only.

The broad Hindu conception of number was, in contrast to the discoveries of the Pythagoreans, extremely useful for computational purposes. To understand the practical value of the Hindu definition

of number, comparing the Hindu concept of number with that of the Greeks is helpful. Recall that the Pythagorean discovery of positive irrational numbers had little effect on computation in part because the positive rational numbers are very closely placed on the positive part of the real line. It is always possible to find a positive rational number that is close enough to any given positive irrational number. In other words it is always possible to calculate exclusively with rational numbers and at the same time maintain a negligible error in the results. There is, however, no positive number of any sort “near” the number -100 . It is not possible to approximate negative numbers with positive ones reasonably. Brahmagupta’s conception of number was not only more abstract; it was also more useful from a computational viewpoint. By incorporating 0 and the negative numbers to form a much larger system of numbers, Brahmagupta and his Hindu contemporaries greatly expanded the concept of solution. They achieved greater freedom in their computations. Later generations of Hindu mathematicians would slightly modify Brahmagupta’s work. They would revisit the problem of doing arithmetic with 0, but most of the Hindu conception of number can be found in Brahmagupta’s book.

7

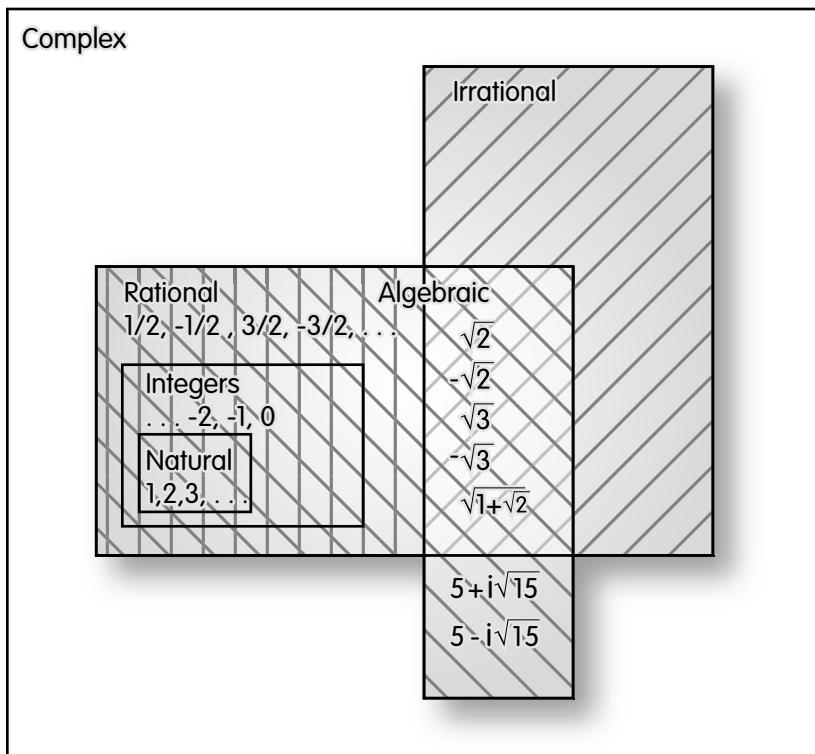
ALGEBRAIC NUMBERS

It is difficult to appreciate how slowly a broad conception of number made its way into the European mathematical tradition. For centuries after the dawn of the Renaissance there were many very prominent European mathematicians who simply refused to accept negative numbers as solutions to equations. The French mathematician and philosopher René Descartes (1596–1650) computed negative solutions to certain equations but dismissed them as “false.” The Scottish mathematician Colin Maclaurin (1698–1746), a very creative and influential mathematician, stated in his work *Treatise of Fluxions* that negative numbers pose a problem to mathematicians in that they give rise to ideas that lacked any physical basis. Despite the widespread aversion to the use of negative numbers, the concept of negative numbers was not—in fact, it could not be—dismissed. Nor was it only negative numbers that these mathematicians had to consider. What we now call complex numbers, numbers that in general have no place on the real line, had been identified early in the 16th century and described as “imaginary.” This new class of numbers could not be dismissed either, however. Both negative and complex numbers arose in the solution of certain common equations. Ideas about equations and the numbers that constituted their solutions evolved together. Each contributed something to the development of the other.

The equations that are most important to the development of the concept of number are called algebraic equations. Algebraic equations are equations of the form $a_nx^n + a_{n-1}x^{n-1} + \dots + a_2x^2 + a_1x + a_0 = 0$, where each x^i is a power of the variable x (so, for example,

when $i = 3$, x^3 means $x \times x \times x$) and where each a_i , called the coefficient of x^i , is the rational number by which x^i is multiplied. (The subscript i means that a_i is the number associated with x^i .) The positive whole number n , which represents the highest power of x appearing in the equation, is called the *degree* of the equation. When n is equal to 2, for example, we have an equation of second degree, also called a quadratic equation.

When we refer to a solution of the algebraic equation $a_nx^n + a_{n-1}x^{n-1} + \dots + a_2x^2 + a_1x + a_0 = 0$ we mean a number that, if substituted for x in the equation, yields a true statement. (Remember that the coefficients of the equation are assumed to be known.) Solutions of algebraic equations are commonly called



During the Renaissance mathematicians discovered new algorithms that enabled them to discover and manipulate new kinds of numbers. Still unanswered, however, was the question of what the new numbers meant.

roots. We can also use the term *root* synonymously with the word *solution*.

Second-degree algebraic equations had been known in various forms since antiquity. Mesopotamian, Greek, Chinese, Hindu, and Islamic mathematicians were already familiar with them, and in each of these cultures mathematicians had also studied some equations of degree higher than 2. The Mesopotamians had solved a very narrow class of fourth-degree algebraic equations, but they could not solve most fourth-degree equations. Hindu mathematicians were adept at solving second-degree equations, as were Islamic mathematicians. Moreover Islamic mathematicians developed efficient numerical algorithms for approximating solutions to equations of the third degree. So the challenge of solving algebraic equations was not a new one even though little progress had been made in finding exact solutions during the thousands of years of recorded history before the Renaissance.

Beginning in Renaissance Italy, however, there was a sudden burst of creativity and insight into the exact solution of equations of third and fourth degree. Never before had exact solutions for general equations of the third and fourth degree been discovered. These new algorithms profoundly affected the course of mathematics. This was not because these mathematicians entirely understood what they had discovered. They did not. In fact centuries passed before they worked out the meaning of their discoveries. Nevertheless the discoveries themselves caused a prolonged burst of mathematical research. The types of numbers that these mathematicians had uncovered when they solved equations of the third and fourth degree were surprising. Even more than surprising, the new numbers challenged their conception of what a number is. The new numbers—although many of the mathematicians who computed them refused to accept them as numbers at all—posed a continuing challenge to every mathematician's understanding. Like a pebble in one's shoe, they were impossible to ignore. Their existence demanded an explanation, but for a very long time no satisfactory explanation was forthcoming. The centuries-long struggle to understand how these new types of numbers fit into a broader number scheme resulted in the number system that we know today.

Numbers that are solutions to algebraic equations are called algebraic numbers. New types of algebraic equations introduced new algebraic numbers. For the next few centuries the historical development of our number system proceeded in tandem with mathematicians' understanding of algebraic equations.

Tartaglia, Ferrari, and Cardano

The discovery of new algebraic numbers begins with the Italian mathematician and physicist Niccolò Fontana (1499–1557), better known as Tartaglia. Tartaglia was party to one of the most famous disputes in the history of mathematics. One reason the dispute is so famous is that it was so crass, but another reason is that the source of the dispute, the book *Ars Magna* by Girolamo Cardano, marks for many historians the beginning of modern mathematics.

Apart from this dispute, about which we have more to say later, Tartaglia led an active and eventful life. He was born in the city of Brescia, an ancient town located in what is now northern Italy. It was once part of the Roman Empire. In Tartaglia's youth Brescia was one of the wealthiest cities in the region, but the young Tartaglia did not share in that wealth. His father, Michele, a postal courier, died when Tartaglia was still a boy and the family was left impoverished. Not much else about Tartaglia's youth is certain. According to one story Tartaglia, at the age of 14, hired a teacher to help him learn to read. According to the



Niccolò Fontana, better known as Tartaglia, discovered a method for finding the roots of any third-degree equation. (Library of Congress, Prints and Photographs Division)

story by the time he reached the letter *k* he ran out of money and from that point on he was self-taught.

In 1512 French forces attacked Brescia and sacked the city. During the battle Tartaglia was severely cut across the face by a sword. The wound left him with a lifelong speech impediment. *Tartaglia* means "stammerer," and for whatever reason, Niccolò Fontana adopted the nickname as his own.

Tartaglia published several books during his life. Some were translations of Greek texts and two were original works. One of his translations was of Euclid's *Elements*. One of his own works, the ballistics treatise *Nova Scientia* (A new science), marks an early attempt to discover the physics of falling bodies. The other book was a three-volume treatise on mathematics, *Trattato di Numeri et Misure* (Treatise on numbers and measures). Today Tartaglia is best remembered for his discovery of a single algorithm. He discovered a method for finding *exact* solutions to algebraic equations of the third degree. His method is, essentially, a formula that uses the coefficients of the equation to compute the roots. Tartaglia's discovery turned out to have many important implications for the history of mathematics, most of which Tartaglia could not have foreseen. He did, however, immediately recognize that he had done something important. (Tartaglia was not the only one to discover a method for finding exact solutions to equations of the third degree. A professor of mathematics at the University of Bologna, Scipione del Ferro (ca. 1465–1526),



Girolamo Cardano. Many scholars believe that the publication of Girolamo Cardano's book Ars Magna marks the beginning of modern mathematics. (Library of Congress, Prints and Photographs Division)

learned to solve a more restricted class of equations of the third degree, but his solution was not widely circulated and he had little impact on the field.)

To appreciate what Tartaglia accomplished, recall that mathematicians in several different cultures and times had all used variants of what we call the quadratic formula to find exact solutions for equations of the second degree. No one in history had found an algorithm for generating exact solutions for equations of the third degree. When Tartaglia announced that he had found such an algorithm, his contemporaries were understandably skeptical. A contest was proposed between Tartaglia and a student of del Ferro named Antonio Maria Fior. Tartaglia and Fior drew up a list of problems for the other to solve. The problems were accompanied by solutions. The solutions were given to a third party, and then Fior and Tartaglia exchanged problems. At the end of the contest Tartaglia had solved all the problems Fior had given him, but Fior had solved none of Tartaglia's. Tartaglia became famous.

Intrigued by news of Tartaglia's discovery, the prominent Italian physician, mathematician, astrologer, and gambler Girolamo Cardano (1501–76) contacted Tartaglia and asked for information about his method. Tartaglia refused. Cardano, however, was not a man who was easily discouraged. (He had, for example, been twice rejected by the College of Physicians in Milan, but he persisted in his attempts to gain admission and was admitted on the third try.) Cardano maintained contact with Tartaglia. He reasoned with him. He badgered him. He tried to bribe him. He was indefatigable. Eventually in a moment of weakness Tartaglia relented and shared the formula with the prominent physician. For his part Cardano agreed never to share the secret with anyone else.

Another important discovery soon followed. Cardano quickly shared Tartaglia's secret with the Italian mathematician Lodovico Ferrari (1522–65). He asked Ferrari to find an algorithm for solving equations of fourth degree, and Ferrari was only too willing to help. Ferrari had been born into a poor family. He initially sought work in Cardano's household as a servant. Cardano, who by this time had become very successful, recognized Ferrari's consider-

able abilities and took him under his wing. Cardano made sure that Ferrari received an outstanding education. In addition to mathematics Ferrari, with Cardano's help, learned Latin and Greek, the two languages essential to scholars of that era. Ferrari never forgot the debt he owed Cardano. He was fiercely loyal to his teacher and mentor, and when he discovered the algorithm for finding exact solutions to arbitrary algebraic equations of the fourth degree, he did not hesitate to share his discovery with Cardano. In an age when scholars practically worshiped the accomplishments of "the ancients," Tartaglia and Ferrari had done something in mathematics that was truly new.

With both algorithms in hand, Cardano published his best-known book, *Ars Magna* (Great art), which included Tartaglia's famous formula as well as that of Ferrari. That book made Cardano famous as a mathematician. As might be expected, the publication of *Ars Magna* infuriated Tartaglia.

What is important for this history is that the algorithms of Tartaglia and Ferrari eventually resulted in a new concept of number. Some of the numbers that arose as solutions to these equations could not be represented as the lengths of line segments; nor could they be used to identify the number of objects in a set. The complicated algorithmic machinations of Tartaglia and Ferrari indicated the existence of new and not-very-intuitive numbers. Even if one did not accept these numbers as "real"—and in his book Cardano calls them "fictitious"—the numbers could be manipulated according to the rules of arithmetic. More importantly they could be used to generate new and valid results.

Cardano half-recognized that these fictitious numbers have a mathematical existence all their own. Cardano was more than devious; he was also a first-rate mathematician. In *Ars Magna* he shows the reader how to manipulate these new numbers to find new solutions to previously unsolvable problems, although he clearly does not believe that these numbers have any practical utility. For example, at one point he shows that he can find two numbers whose sum is 10 and whose product is 40. It is not difficult to show that if we restrict our attention to numbers that lie on the real line, this problem has no solution. The product of any two *real*

numbers that add up to 10 can be no larger than 25. Cardano, however, realizes that this problem has solutions that are not real. He produces the pair of numbers $5 + \sqrt{-15}$ and $5 - \sqrt{-15}$ as solutions to his problem. These numbers are the type of numbers that had become familiar to him through his study of algebraic equations. Notice that there is no real number whose square is -15 , so neither $5 + \sqrt{-15}$ nor $5 - \sqrt{-15}$ can be a real number. However, if one multiplies these two numbers together one finds that their product is 40, just as Cardano claimed, and their sum is clearly 10 because the terms involving $\sqrt{-15}$ and $-\sqrt{-15}$ cancel with each other. Cardano is clearly proud of this demonstration, but he cannot get himself to embrace these new numbers. In *Ars Magna* he says, “So progresses arithmetic subtlety the end of which, as is said, is as refined as it is useless.”

Part of the difficulty in expanding the set of numbers to include all algebraic numbers is that the extension of the old concept of number required to accomplish this is huge. The jump from the set of positive numbers with which these mathematicians were familiar to the set of algebraic numbers required that they solve several problems simultaneously. We previously mentioned that the set of positive numbers is closed under addition, multiplication, division, and the extraction of roots to the extent that every positive number has a positive root. Initially it may seem that the only problem is the operation of subtraction, because the positive numbers are not closed under subtraction. Simply introducing the negative numbers, however, introduces a new problem. The set of all real numbers is closed under addition, subtraction, multiplication, and division. It is not closed under the extraction of roots: For example, square roots of negative numbers are never real numbers. Introducing the negative numbers solves one set of problems but simultaneously creates another.

For a long time after the discoveries of Tartaglia and Ferrari there was considerable disagreement about which numbers were “real” numbers—real in the sense that they had some foundation in reality—and which numbers were “fictitious” or “imaginary,” that is, those numbers that had no basis in reality. Many mathematicians were understandably reluctant to spend their time studying or

using numbers that they considered meaningless. Brahmagupta's bold idea of effectively expanding the definition of number to include all numbers that arose in the course of a calculation was not an idea that European mathematicians readily accepted. To use any number generated by a calculation raises—but does not answer—the question, What do these new numbers mean?

Despite these difficulties many mathematicians in the era following Cardano recognized the need to expand their concept of number to include more of the solutions their algorithms were generating. They recognized that without a better, more inclusive definition of number, they could not develop a deep understanding of the algebraic equations that many of them were studying.

Girard and Wallis

The Flemish mathematician Albert Girard (1595–1632) was one of the first mathematicians to adopt a conception of number broad enough to serve as a help rather than a hindrance in the analysis of algebraic equations. Not much is known of Girard as an individual. Different biographies even give slightly different dates for his birth and death. In his own time he was generally described as an engineer rather than as a mathematician, but this is not especially surprising. He had skill in engineering, a talent with which many people can identify, and he was far-sighted in mathematics, a talent with which most people cannot identify. In any case in 1629 in his work *Invention nouvelle en l'algèbre* Girard proposes an extension of the concept of number that far surpasses that of his contemporaries.

First, his understanding of negative numbers is modern. He conceives of negative numbers as directed distances along a line, a clear anticipation of the real number line, and in his solutions to algebraic equations he uses positive, negative, and complex numbers. Nor does he use them only in the course of solving these problems. Unlike his contemporaries, he accepts all of these numbers as legitimate roots to algebraic equations. This expanded concept of number allowed him great freedom in the study of algebraic equations.



John Wallis, one of the most influential mathematics of his time, was remarkably forward-thinking in his treatment of complex numbers.
 (Science Museum, London/Topham-HIP/The Image Works)

Second, Girard was interested in the relationship between an algebraic equation and its roots or solutions. Usually we are given an algebraic equation and our goal is to find the roots, but it is possible to work the problem in reverse: Given a collection of numbers can we find an algebraic equation with these numbers as roots? The answer is yes, and Girard made some headway in identifying the relationships that exist between the roots of an algebraic equation and the coefficients that appear in the equation itself. (By contrast if one refuses to accept negative and complex roots then it is impossible to see the relationships that exist between the equation and its roots.)

In the end Girard draws an extraordinary conclusion: He speculates that an algebraic equation has as many roots (solutions) as the degree of the equation. (The *degree* of an algebraic equation is the largest exponent appearing in the equation.) He is right, provided one counts negative numbers and complex numbers as valid roots. This is an important insight into the relationship between algebraic equations and algebraic numbers. Unfortunately he cannot prove that his conclusion is correct. Girard's speculation is not proved until centuries after his death. In retrospect this is not surprising. The name of the result is the fundamental theorem of algebra, and a great deal more mathematics had to be developed before proving the fundamental theorem became possible. Girard's ideas drew little attention at the time he proposed them.

By way of contrast with the ideas of Girard, the French mathematician and philosopher René Descartes's (1595–1650) *La Géométrie* asserts that an algebraic equation can have no more roots than the degree of the equation. This is a much more conservative statement. Descartes's ideas arose out of his unwillingness to expand his concept of number. He was willing to accept fewer roots than the degree of the equation because he could not accept negative and complex roots as valid solutions.

Another prominent mathematician who exhibited insight into numbers was the British mathematician John Wallis (1616–1703). Wallis discovered mathematics fairly late in life. English schools did not emphasize mathematics except as an aid for commerce and the trades. Because Wallis had an upper-class education he did not encounter mathematics before college. Nor did he encounter mathematics in college. He began his adult life as an ordained minister. During this time Britain was engaged in a civil war. Wallis supported the group led by Oliver Cromwell, and when Cromwell's forces captured enemy messages written in code Wallis quickly deciphered them. He earned Cromwell's favor for this accomplishment. A few years later Wallis read his first book on higher mathematics, *The Keys to Mathematics* by William Oughtred. Wallis discovered that he had a talent and a love for the subject, and he tenaciously researched and published his mathematical ideas for the rest of his life.

Wallis proved to be a very creative mathematician. He made contributions to algebra and the branch of mathematics that would soon be called calculus. Isaac Newton claimed that he was inspired to study mathematics after reading Wallis's *Arithmetica Infinitorum*. For the last several decades of his life Wallis had an influence felt throughout the European mathematical community.

In the history of numbers Wallis made an interesting and farsighted contribution. Wallis became interested in geometrically representing the roots of quadratic, or second-degree, algebraic equations. He accepted the idea that algebraic equations have negative and complex solutions. He accepted the idea of representing negative numbers as directed distances along a line. In other words he imagined the positive numbers on one side of 0 and the nega-

tive numbers, their mirror image, placed on the other side. The question, then, was where to place numbers that are complex; that is, how could he geometrically interpret numbers that are not positive, negative, or 0?

Wallis imagined the complex numbers that arose as solutions to quadratic equations as the dimensions of squares with negative areas. Since $\sqrt{-1}$ has no place on the real line—there is no real number with the property that if we multiply it by itself the answer is -1 —these complex numbers could only be represented in two dimensions. In the number that we would write as $a + b\sqrt{-1}$, Wallis imagined a as a directed distance along what we call the real number line. He imagined b as a directed distance measured in a direction perpendicular to the real number line at the point a . The only role of the $\sqrt{-1}$ is to indicate that b and not a lies off the real line. In this way Wallis associated a picture with each number of the form $a + b\sqrt{-1}$. This was a very insightful way of describing complex numbers, because it associated a plausible interpretation with what most mathematicians of the time considered nonsense. Wallis's conception was as modern as any conception of complex numbers for almost a century, but it did not have much effect on the ideas of his contemporaries.

Wallis's conception of the number system was as forward-looking as any for the next several decades. The German mathematician and philosopher Gottfried Leibniz, one of the preeminent mathematicians of his day, showed a great deal of facility in manipulating complex numbers, but despite his skill, he remained unconvinced about their reality. This was the general attitude toward negative and complex numbers for a very long time. Mathematicians continued to gain expertise in working with the “new” numbers. They found their newly acquired skill useful, because it enabled them to obtain results that they otherwise could not. At the same time most mathematicians worked hard to avoid negative and complex numbers in their answers, because the interpretation that they placed on these numbers caused them to be suspicious of any quantity that they represented. As late as the first half of the 19th century, the prominent and prolific

English mathematician Augustus De Morgan wrote, “We have shown the symbol $\sqrt{-a}$ to be void of meaning, or rather self-contradictory and absurd. Nevertheless, by means of such symbols, a part of algebra is established which is of great utility.”

Although De Morgan was one of the major mathematicians of his day, he was behind the times in his conception of complex and negative numbers. The tide had begun to turn even before his birth. The Swiss mathematician Leonhard Euler (1707–83) had discovered many important relations among positive, negative, and complex numbers and had learned to incorporate all of these numbers into a theory of functions that produced many important results.

Euler and d'Alembert

Leonhard Euler, the most prolific mathematician in history, originally planned to enter the ministry. His father, Paul, was a minister, and his mother, Margaret, was the daughter of a minister. In addition to being a minister, Paul Euler was something of a mathematician and he helped his son develop his mathematical talent, though apparently both parents hoped that their son would enter the ministry. As is any good son Leonhard Euler was obedient to a point. At the University of Basel he studied theology as well as medicine, languages, mathematics, physics, and astronomy. He took classes from Jacob Bernoulli, one of the great mathematicians of the day, and was a friend of Jacob's sons, Nicolaus II and Daniel Bernoulli. (The Bernoulli family produced many prominent mathematicians, including Jacob, Nicolaus, and Daniel.) Eventually the three young men, Nicolaus, Daniel, and Leonhard, found work at the Saint Petersburg Academy in Saint Petersburg, Russia.

Euler had an extraordinary capacity for mathematics. He wrote on almost every branch of mathematics: groundbreaking research and elementary mathematics books alike. He seemed equally at home with both ends of the mathematical spectrum. It is said that he could perform enormous calculations in his head with little difficulty and that he could sit and write mathematics for prolonged periods without pausing, much as Mozart is said to have

written music. Even blindness—and Euler was blind for the last 17 years of his life—had little effect on his output. He continued to dictate mathematics papers right to the end. In the 20th century Euler's collected works, including correspondence, were published by Birkhäuser Publishing of Basel, Switzerland. It is an 82-volume set.

THE DEBATE OVER “FICTITIOUS” NUMBERS

What follows are excerpts from the 17th-century text *A Treatise on Algebra* by John Wallis. Keep in mind that Wallis was one of the foremost mathematicians of his time. Notice how he struggles to communicate his ideas about the utility of negative and complex numbers as roots for quadratic equations, the same equations and the same numbers that many students now encounter in high school:

These imaginary quantities (as they are commonly called) arising from the supposed root of a negative square, (when they happen,) are reputed to imply that the case proposed is impossible.

And so indeed it is, as to the first and strict notion of what is proposed. For it is not possible, that any number (negative or affirmative) multiplied into itself, can produce (for instance) -4 . Since that like signs (whether + or -) will produce +; and therefore not -4 .

But it is also impossible, that any quantity (though not a supposed square) can be negative. Since that it is not possible that any magnitude can be less than nothing, or any number fewer than none.

Yet is not that supposition (of negative quantities,) either unuseful or absurd; when rightly understood. And though, as to the bare algebraic notation, it import a quantity less than nothing: yet, when it comes to a physical application, it denotes as real a quantity as if the sign were +; but to be interpreted in a contrary sense.

As for instance: supposing a man to have advanced or moved forward, (from A to B,) 5 yards; and then to retreat (from B to C) 2 yards: if it be asked how much he had advanced (upon the whole march) when at C? or how many yards he is

In the field of complex numbers Euler invented much of the symbolism that we use today and he discovered many important relations. He was the first to use the letter i for $\sqrt{-1}$, the notation that is now standard. Euler studied logarithms of negative and complex numbers. He studied expressions such as $(a + bi)^{c+di}$. The number with the decimal expansion 2.71828 . . . was also studied

now forwarder than when he was at A? I find (by subducting 2 from 5,) that he is advanced 3 yards. (Because $= 5 - 2 = +3$.)

But if, having advanced 5 yards to B, he thence retreat 8 yards to D; and it then asked, how much he is advanced when at D, or how much forwarder than when he was at A: I say -3 yards. (Because $5 - 8 = -3$.) That is to say, he is advanced 3 yards less than nothing. . . .

Now what is admitted in lines, must on the same reason, be allowed in plains also.

As for instance: supposing that in one place, we gain from the sea, 30 acres, but lose in another place 20 acres: if it be now asked, how many acres we have gained upon the whole: the answer is, 10 acres or $+10$. (Because $30 - 20 = 10$.) (For the English acre being equal to a plain of 40 perches in length, and 4 in breadth, whose area is 160; 10 acres will be 1600 square perches.) . . .

but if then . . . we lose 20 acres more; and the same question be again asked, how much we have gained in the whole; the answer must be -10 acres. (Because $30 - 20 - 20 = -10$.) That is to say, the gain is 10 acres less than nothing. Which is the same as to say, there is a loss of 10 acres. . . .

And hitherto, there is no new difficulty arising, nor any other impossibility than what we met with before, (in supposing a negative quantity, or somewhat less than nothing:)

But now (supposing this negative plain, -1600 perches, to be in the form of a square:) must not this supposed square be supposed to have a side? And if so, what shall this side be?

We cannot say it is 40, nor that it is -40 . (Because either of these multiplied into itself, will make $+1600$; not -1600 .)

But thus rather, that it is $\sqrt{-1600}$, (the supposed root of a negative square;)

(*Wallis, John. Treatise on Algebra. London: J. Playford, 1685.*)



Leonhard Euler, one of the most prolific mathematicians in history
(Library of Congress, Prints and Photographs Division)

by Euler, and he gave it the name e . This notation has also become standard. (The number e is very important in calculus.) Euler discovered the remarkable equation $e^{i\pi} - 1 = 0$, a result that links the so-called five most important numbers in mathematics in a single equation. All of these discoveries and more require deep insight into what a number is, but Euler's abstract manipulations might also have been dismissed as abstract nonsense had not Euler also begun to discover new and important uses for these numbers.

Euler greatly increased the importance of complex num-

bers in mathematics. Some of his insights into complex numbers were of a more technical nature: He learned how to extend the definition of many functions that had once been defined only for real numbers so that they were also defined for arbitrary complex numbers. This made these functions more useful. Some of his insights were more down-to-earth. He was able, for example, to use complex numbers in the study of cartography. This is not to say that he was always right when he wrote about complex numbers. He made occasional errors. Nor was he entirely free of the unease that his contemporaries so frequently expressed about the “absurdity” of these numbers. Nevertheless he saw deeply into the nature of complex numbers and greatly expanded their prominence in mathematics.

There was a time in Euler's life when he was friends with the French mathematician Jean Le Rond d'Alembert (1717–83), and it is d'Alembert who made the next important discovery about com-

plex numbers. (Euler and d'Alembert later had a falling out.) D'Alembert was a prominent and controversial figure in the world of mathematics. His biological parents were wealthy and socially prominent, but they had no interest in raising d'Alembert. Instead they placed their infant son with a family of modest means by the name of Rousseau. D'Alembert was raised by the Rousseaus, but he maintained some contact with his biological father. It was from his biological father that he received the money needed to pay for his excellent education. D'Alembert attended Collège des Quatre Nations, where, as Euler did, he studied a little of everything: mathematics, science, medicine, and law. This solid and very broad education was a good investment. Although success occurred early for d'Alembert—he was a member of the Académie des Sciences at the age of 24—he continued to live with the Rousseaus until he was middle-aged.

Today d'Alembert is remembered for several diverse accomplishments. First, he was an insightful mathematician who had a particular interest in the mathematical consequences of Newton's three laws of motion. Second, he is remembered as the editor of an influential encyclopedia. Nor were his duties confined to editing the encyclopedia. He also wrote most of the science articles, a massive undertaking. He is also remembered for an insightful article that he wrote about mathematics and public health with special reference to the disease smallpox.

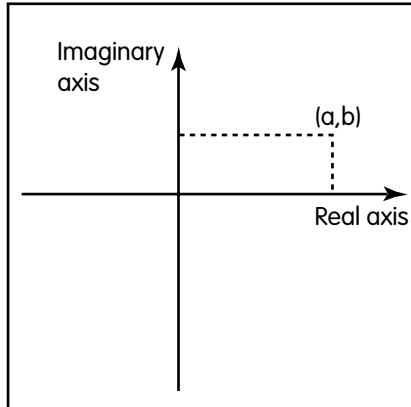
Of particular interest in this history d'Alembert made an important contribution to our understanding of the number system. Essentially d'Alembert wanted to know whether the set of numbers of the form $a + bi$, where a and b represent real numbers, is "big enough" to include all solutions to all algebraic equations. For example, is the set of numbers of the form $a + bi$ closed under the basic arithmetic operations: addition, subtraction, multiplication, division, and the extraction of roots? More generally he wanted to know whether any combination of these operations when applied to a number of the form $a + bi$ produces another number of *this same form*. For example if we take the cube root of $2 - 3i$, is our answer also of the form $a + bi$? The answer, which is yes, is not at all obvious.

This was an ambitious goal. D'Alembert was, in some ways, attempting to conclude thousands of years of research and speculation about algebraic numbers. The problem had begun more than 2,000 years earlier when the ancient Greeks discovered that the square root of a positive rational number is not necessarily a rational number. For the next two millennia mathematicians had worked to construct a number system big enough to be closed under all the operations of arithmetic and the extraction of roots. D'Alembert wanted to show that if he added, subtracted, multiplied, divided, or extracted a root from a number of the form $a + bi$ the result was another number of the form $a + bi$. He concluded that, in fact, it was. He concluded that the set of all complex numbers is big enough to be closed under these ancient operations.

D'Alembert almost got the proof right. His proof contained some small flaws but these were quickly identified and corrected. Furthermore Euler was able to extend this result to show that many other functions of interest to mathematicians—the logarithmic functions, the trigonometric functions, the exponential function, and so forth—when applied to complex numbers yield numbers that can also be written in the form $a + bi$. In fact the expression $a + bi$ is one way of defining what a complex number is, so another way of describing d'Alembert's and Euler's discoveries is to say that they show that the complex numbers are closed under the operations that they considered. The complex number system turned out to be the minimal extension of the set of positive rational numbers that is closed under any combination of the usual mathematical operations. The problem that had first plagued the Greeks in the year 600 B.C.E was solved in the 18th century.

It may seem that complex numbers were finally understood, but there was more to be done. Mathematicians were slowly developing a clearer understanding of how real numbers can be geometrically represented as directed distances on a line, but the question remained: How can complex numbers be geometrically represented? The answer really lies in the fact that every complex number is of the form $a + bi$, so that every complex number has a “real” part,

which we represent with the letter a , and an imaginary part, the “ b part.” (Here the letter i serves only to distinguish the “real” part of the number from the imaginary part; the imaginary part is the coefficient of the i term.) Notice that when b is 0, the number $a + bi$ is a real number. This is an important observation because the real part of the complex number can be represented as a directed distance on the real number line. Likewise the imaginary part can be represented as a directed distance on an “imaginary number line.” The answer to the question of how to represent the set of all complex numbers geometrically, then, consisted of setting up a correspondence between the set of numbers $a + bi$ and the points in the plane, which can be represented by ordered pairs of numbers of the form (a, b) (see the above diagram).



The Argand diagram represents complex numbers geometrically by establishing a one-to-one correspondence between complex numbers and points of a plane.

The Complex Numbers: A Modern View

The first person to have the idea of representing complex numbers geometrically was Caspar Wessel (1745–1818), a Norwegian mathematician, surveyor, and mapmaker. Wessel spent most of his life surveying and making maps, and his discovery, which was published in 1799, appeared as part of a paper on surveying. It attracted no notice from mathematicians until almost 100 years later, when it was rediscovered and republished for the benefit of mathematicians. Wessel’s idea, however, had already been rediscovered more than once in the intervening years. In 1806 Jean Robert Argand (1768–1822), a French bookkeeper, accountant, and mathematician, rediscovered the geometrical interpretation of

USING COMPLEX NUMBERS

What we now call complex numbers were initially viewed with a great deal of skepticism. The terminology in use today continues to reflect the old view that complex numbers have no basis in reality. Even now when mathematicians refer to a complex number written in the form $a + bi$, they continue to use the old terminology: The letter a represents the *real* part of the number, and the letter b is the *imaginary* part. The square root of a negative number has no real part and so it is often described as a pure imaginary. It is easy, therefore, to conclude that complex numbers have no applications outside mathematics, but nevertheless they do. In fact the study of functions of a complex variable is an important part of the education of many engineers and scientists. Here are three common applications:

1. Fluid dynamics: The study of fluid motions—and here the word *fluid* refers to gases as well as liquids—is one of the oldest and most useful branches of all of science. It plays an essential role in the design of aircraft, ships, and certain biomedical devices. The use of complex numbers in the study of two-dimensional flows is a part of the science, which is about 100 years old, and it is still finding new applications. One of the earliest practical uses for functions of a complex variable in fluid dynamics was in the study of aircraft wing design. The field of complex variables provides a quick and reasonably accurate way for engineers to model the flow of a gas or liquid

complex numbers. His name is attached to the usual geometric interpretation of complex numbers, called the Argand diagram. The discovery was made a third time in 1831 by the German mathematician Carl Friedrich Gauss (1777–1855), one of the most prolific mathematicians in history. The Hungarian mathematician János Bolyai (1802–60), one of the founders of non-Euclidean geometry, also discovered something similar. This was clearly an idea whose time had arrived.

Though the idea of representing complex numbers as points on a plane is simple, it is very important to those mathematicians concerned with studying functions of a complex variable. Its impor-

over a surface. It is a mathematical method that is learned by all undergraduates studying aerospace engineering.

2. Differential equations: The field of differential equations is one of the most important of all branches of mathematics, because the laws of nature are generally expressed as differential equations. In a differential equation the function itself is the unknown. The equation can tell us how the function changes over time, or, perhaps, how the function changes from one point in space to another. The goal of the mathematician is to discover as much as possible about the unknown function from its differential equation. Methods of solution for these equations often involve functions of one or several complex variables.
3. Heat conduction: Here the goal is often to compute the temperature of the interior of a body when the only measurements we can make are at the surface. This is important whenever we are trying to predict the reaction of a body to the transfer of heat into or out of it, an important engineering problem. Once again techniques that involve functions of a complex variable play an important role in helping engineers solve these types of equations. These techniques are firmly established—they became important almost as soon as scientists convinced themselves that complex numbers might be useful and continue to play an important role in this branch of engineering today.

tance lies in the fact that a geometric representation of a complex number is a conceptual aid in a subject that even at its simplest is still quite abstract. For example, the distance of any real number to the origin of the real number line is the *absolute value* of the number. The numbers 3 and -3 are both located a distance of three units from the origin. This idea is expressed with the equation $| -3 | = | 3 | = 3$, where the vertical lines represent the absolute value. What, however, is the distance from the origin to the number $2 + 3i$? Once the number $2 + 3i$ is represented as the point $(2, -3)$, the answer is visually obvious. The method of computing the distance is to use the Pythagorean theorem: $| 2 + 3i | = \sqrt{2^2 + 3^2}$.

The advantages of using the complex plane, however, extend far past computing absolute values. To appreciate the difficulty of using complex numbers before the invention of the Argand diagram, consider the problem of graphing functions. In the study of functions of a real variable, every student learns to associate a two-dimensional graph with the algebraic description of the function of interest. The graph is important because it establishes a relationship between a geometric and an algebraic interpretation of the function. The picture helps us understand the function better. By contrast the study of functions of a complex variable *begins* at a much higher level of abstraction. The Argand diagram uses two dimensions to represent a single independent complex variable and another two dimensions to represent the dependent complex variable. As a consequence the simplest graphs of functions of a complex variable are four-dimensional. At this level of abstraction any conceptual tool can make a big difference. In fact it was not until mathematicians were able to represent complex numbers as points on a plane that the study of functions of a complex variable really flourished.

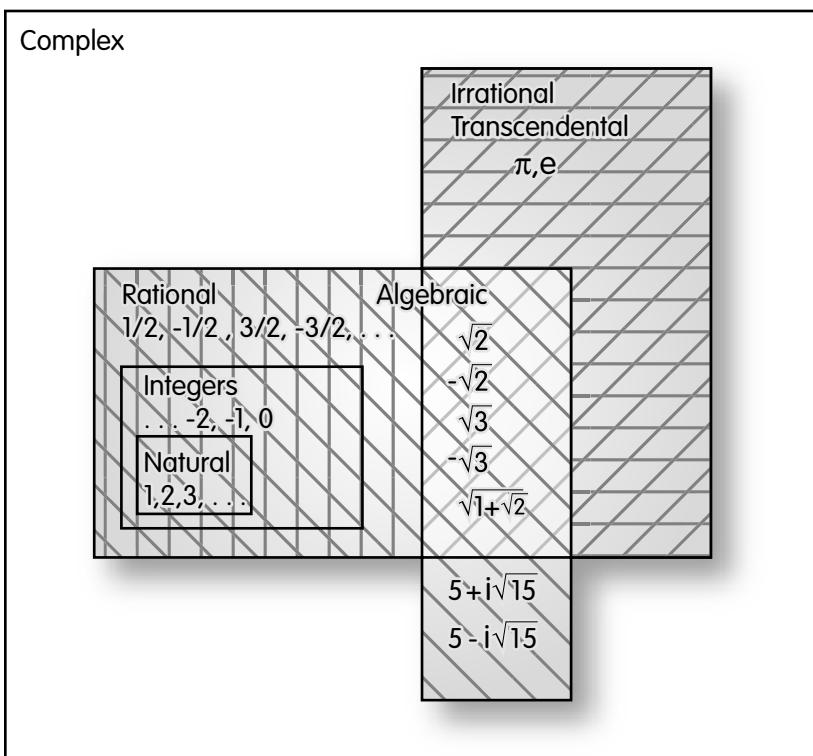
8

TRANSCENDENTAL NUMBERS AND THE SEARCH FOR MEANING

The study of algebraic equations helped mathematicians learn about new kinds of numbers. (Recall that an algebraic equation is an equation of the form $a_nx^n + a_{n-1}x^{n-1} + \dots + a_2x^2 + a_1x + a_0 = 0$, where each x^i is a power of the variable x —so, for example, when $i = 3$, x^3 means $x \times x \times x$ —and where each a_i , called the coefficient of x^i , is the rational number by which x^i is multiplied. The subscript i indicates that a_i is the number that multiplies x^i .) Every root, or solution, of an algebraic equation is called an algebraic number. There are infinitely many algebraic numbers. Every rational number is an algebraic number because every rational number, whether it is positive or negative, is the root of an equation of the form $x + a_0 = 0$. Many irrational numbers are algebraic as well. It may seem that every number is algebraic, but this is not true. It is a surprising fact, about which we have more to say later, that *most* numbers are *not* algebraic. As one might expect, mathematicians were quick to give a name to these nonalgebraic numbers. Perhaps surprisingly the name was chosen before it was shown that any such numbers existed! Numbers that are not algebraic are called transcendental. A *transcendental number* is not the root of any algebraic equation.

Euler suspected that transcendental numbers exist, though he was unable to prove that any particular number is transcendental. Of particular interest to Euler was the number that mathematicians call e . The number e is especially important in calculus

and the branches of mathematics that grew out of calculus. Because the number $e = 2.718 \dots$ has an important role in many mathematical and scientific problems, Euler spent a fair amount of effort studying the properties of e . He knew that e is irrational, but he could not show that it is transcendental; that is a much more difficult problem. The difficulty lies in the fact that to show that e or any other number is transcendental, one must show that there is no algebraic equation that has e as a root. In other words one must show that no matter how we choose the degree of the equation, and no matter how we choose the rational coefficients that appear in the equation, when we substitute e for x in the expression $a_nx^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$ the result is



The existence of transcendental numbers was finally established in the second half of the 19th century. The numbers π and e are the two best-known transcendental numbers, but many others exist.

never 0. This is difficult for any transcendental number. As Euler did, most mathematicians interested in transcendental numbers put their effort into trying to establish the transcendental nature of the two numbers e and π .

The Swiss mathematician, philosopher, and astronomer Johann Heinrich Lambert (1728–77) was one of the first to gain some insight into the nature of the number π . Lambert must have been an interesting man. He was a member of a family of modest means—his father was a tailor—and from a young age he studied geometry and astronomy on his own. He studied astronomy with instruments that he designed and constructed himself. In fact he was self-taught in all of the subjects in which he excelled. As an adult he earned a living as a bookkeeper and later as a tutor, but his primary interest was always scholarly work. He must have been a very determined scholar. He received recognition for his efforts fairly late in life. In 1768 Lambert published the mathematical result for which he is best known: He showed that π is irrational. This was an important first step.

Toward the end of the 18th century the French mathematician Adrien-Marie Legendre (or Le Gendre; 1752–1833) speculated that π is transcendental, but again neither he nor anyone else at the time could prove that. It was not until about 50 years later that the French mathematician Joseph Liouville (1809–82) proved the existence of transcendental numbers by taking an entirely different tack.

Liouville spent most of his life in Paris teaching and researching mathematics. He showed mathematical promise at a very early age. By the time he was a teenager he was writing papers for mathematical journals. He contributed to many different areas of mathematics. What is important to the history of numbers, however, is that Liouville discovered a set consisting of infinitely many transcendental numbers. All of the numbers that Liouville discovered have the same form. They are all constructed according to the same general formula. We omit his somewhat complicated formula and instead give an example of a number constructed according to Liouville's formula: 0.1100010000000000000000100 . . . where we place a 1 in the first place to the right of the decimal point, a 1 in the second place ($2 = 2 \times 1$), a 1 in the sixth place

($6 = 3 \times 2 \times 1$), a 1 in the 24th place ($24 = 4 \times 3 \times 2 \times 1$), and so on. Everywhere else we write a 0.

In 1873 almost 30 years after Liouville proved the existence of a set of transcendental numbers, the French mathematician Charles Hermite (1822–1901), after a great deal of work, completed Euler's project and proved that the number e is transcendental. Nine years after Hermite's discovery the German mathematician Ferdinand von Lindemann (1852–1939) demonstrated that the number π is transcendental.

Part of the difficulty with understanding and using transcendental numbers is that there is no easy way to express them. We can simply write a rational number as a fraction with a whole number in the numerator and another in the denominator. We can also write the algebraic equation associated with an algebraic number that is not rational. This is essentially what we do when we write $\sqrt{2}$. This notation indicates that this number is a root of the equation $x^2 - 2 = 0$. However, we can use neither method with transcendental numbers. Transcendental numbers are, as a matter of definition, *not* rational and *not* the root of any algebraic equation. Our method of place-value notation is not especially valuable for describing most transcendental numbers, either. If we write out the number as a decimal, we find that the sequence of digits never ends and never forms a repeating pattern in the way that decimal expansions of rational numbers always do. That is why when we refer to the transcendental number formed by the ratio of the circumference of a circle to its diameter—the number that is approximately 3.14—we use the Greek letter π . Similarly the number that occupies a special place in calculus—its value is approximately 2.71—is identified by the letter e rather than by its numerical value. The same shortcoming in our number system explains why we had to describe a procedure for writing Liouville's transcendental number rather than simply writing the number itself. There is no convenient notation for any of these numbers. This is why, although most real numbers turn out to be transcendental, few people, even few mathematicians, can name more than four or five.

It may seem that with the discovery of transcendental numbers mathematicians finally had a firm grasp of the concept of number. They knew something about rational and irrational numbers, positive and negative numbers, real and complex numbers, the number 0, and algebraic and transcendental numbers. They had found a number system, the complex number system, that is closed under the operations of arithmetic, the extraction of roots, and other more complicated operations. It may appear that they had a firm grasp of the concept of number, but they did not. Furthermore they knew they did not.

For most of the 19th century mathematicians had worked to understand the nature of various classes of numbers, but there was no unity to their discoveries. As biologists who seek to classify the species present in an unfamiliar ecosystem and to identify the relationships between different species do, mathematicians had developed a classification scheme for numbers. Each number could be classified as algebraic or transcendental; irrational or rational; positive, negative, or 0; real or complex. But there was no explanation of how, for example, irrational numbers are related to the rational ones. The only definition in use at the time, a definition going back to the Pythagoreans, was that the irrational numbers are not rational, and this was far from satisfying. A deeper insight into numbers was necessary to unify the subject and to ensure a strong logical foundation on which to build the new mathematics of the time. The German mathematician Richard Dedekind (1831–1916) was the first to introduce a unifying concept to the idea of number.

Dedekind and the Real Number Line

Richard Dedekind was a quiet man. He lived with his sister for much of his adult life, and for 50 years he taught high school mathematics. He received a world-class education in higher mathematics at Göttingen University, one of the major centers of mathematics in the world. He entered Göttingen at 19 years of age and received a doctorate three years later. While working as a high school teacher, he made several important and highly original con-



The Göttingen University Department of Mathematics during what the university calls “the Hilbert era.” As were many of the best mathematicians of his age, Richard Dedekind was educated here. (Courtesy of the University of Göttingen)

tributions to mathematics. His ideas make up an integral part of the education of every mathematician today, but in his time his discoveries received less recognition than they deserved.

Dedekind spent most of the second half of the 19th century teaching high school; during that time mathematicians were discovering many aspects of mathematics that were counterintuitive. They discovered properties of infinite sets of numbers that defied their commonsense notions of what should be true, and they were discovering functions that had strange and unexpected mathematical properties. After one such demonstration Georg Cantor, one of the leading mathematicians of the time, said that although he understood the math he still could not believe the outcome. During the second half of the 19th century it became clear that the old ideas and the old techniques were just not powerful enough for the new math. There was a general recognition that the logical underpinnings of the new mathematics were weak. A

weak foundation could not be expected to support all of the new mathematics that these mathematicians were attempting to create. Mathematics is a deductive discipline. New results are discovered as logical consequences of previously established ones. If the logic of the existing ideas is weak, the deductions may well turn out to be wrong.

It was while he was teaching high school calculus that Richard Dedekind became aware that the definition of number that mathematicians of that time were using was not adequate for what they were attempting to do. He recognized that his contemporaries did not have a firm grasp of the basic properties of the number system. Mathematicians were using numbers without a detailed understanding of their logical structure. Numbers are among the most basic objects of mathematics, and as mathematicians worked to develop an increasingly abstract, logically rigorous subject, they *needed* a clearer understanding of number. Even as late as the 19th century the relationship between rational and irrational numbers was not obvious. Twenty-five centuries after Pythagoras, mathematicians still understood irrational numbers by what they are not. They knew that irrational numbers are not rational. They knew that the decimal expansions of irrational numbers do not terminate and do not follow a repeating pattern. (The decimal expansion of every rational number eventually “settles down” into a repeating pattern.) Knowing only what something is not, however, does not produce much insight. Insights of the time were not adequate for the type of mathematics that was developing in the 19th century. Dedekind set out to correct this deficiency.

His solution was what we call the real number line. We have used the real number line throughout this book as an aid to conveying the development of the concept of number, but the rigorous definition of the real number line was Dedekind’s invention. For Dedekind the line was just the conceptual tool necessary to represent the real number system.

It is an axiom of geometry that a line is continuous: That is, it has no breaks or holes in it. It was from this axiom that Dedekind began to reason. His goal was to show that the number system is

also continuous, and it is continuous in the same way that the line is continuous. In the real number system there are no breaks or holes. There are no gaps between numbers. To show this he needed to find a logical way of establishing a correspondence between points on the line and the system of “real,” not complex, numbers. The key is to imagine a line with the position of the rational numbers marked out on it. Now imagine cutting the line. Every cut divides the line into two sets of numbers. One set of numbers lies to the left of the cut. The other set of numbers lies to the right of the cut. Every rational number in the set on the left is less than every rational number in the set on the right, and every rational number in the set on the right is greater than every rational number in the set on the left. Notice that Dedekind’s cut has divided the line into two mutually exclusive sets.

Next let the letter a represent the position on the line where the cut occurs. The number a is the directed distance from the origin, so a is negative if the cut occurs to the left of the origin and positive if the cut occurs to the right. The number a must be either rational or irrational, because every number is either rational or irrational. The division of the line is completed once a decision has been made about whether a is placed in the right or the left set. Suppose a is assigned to the set on the left. *If a is rational* then the set on the left has a largest rational number. That largest number is a . It is also true that in this circumstance the set on the right does not have a smallest rational number because between any two rational numbers there is always a third, so that no matter which rational number b is chosen from the set on the right there is still a third rational number that is greater than a and less than b . If, however, a is assigned to the set on the left, and if that set still has no largest rational number, then a is irrational. This thought experiment, called a Dedekind cut, was the first to produce a logically complete description of the real number system. It establishes a correspondence between discrete numbers and continuous lines. (Similar statements can be obtained if we assign a to the set on the right. In this case the set on the right contains a smallest rational number if a is rational, and it does not contain a smallest rational element if a is irrational.)

The idea of associating a number with each cut has very important implications for mathematics. It shows that the number system varies continuously, and that there are no gaps. This may seem to be common sense, but it is not logically obvious, and logic, not common sense, is the currency of mathematics. In many ways Dedekind's idea represents the completion of the real number system. After thousands of years of struggling with incomplete number systems and confusing concepts of number, Dedekind had developed a model for the real number system that addressed the confusion that had surrounded numbers since the Pythagoreans uncovered the existence of irrational numbers some 25 centuries earlier. The complex number system, then, can be imagined as the set of coordinates on a plane, where both the real and imaginary axes are completed in the manner of Dedekind. Another consequence of Dedekind's thought experiment is that it allows a positive answer to questions such as, What is $\sqrt{2}$? The answer is $\sqrt{2}$ is the position of the cut such that every rational number, a/b , on the left segment has the property that $a^2/b^2 < 2$, and every rational number, a/b , on the right segment has the property that $a^2/b^2 > 2$. It was Dedekind's ideas that permitted mathematicians to move forward in the study of mathematics, confident that their deductions involving numbers were based on a strong logical foundation.

PART THREE

THE PROBLEM OF INFINITY

9

EARLY INSIGHTS

There is nothing in the world around us that prepares us for the mathematical concept of infinity. Everything around us is finite. All of the best science indicates that the universe had a beginning no more than 20 billion years ago, and, one way or another, the universe will end, either quietly or explosively, sometime in the distant future. The universe, as do all of its occupants, has a finite life span. Furthermore the universe, as big as it is, contains only a



There is no physical object that corresponds to the mathematical idea of infinity. (Courtesy of the National Aeronautics and Space Administration)

finite amount of mass. Scientists have begun to estimate the amount of mass in the universe. There is still a lot of uncertainty in these measurements, but there is no reason to suspect that they are “infinitely” far off: That is to say, there is no reason to believe that there is an infinite amount of mass in the universe.

Although there is no physical analog for the mathematically infinite, no concept could be more important to mathematics. The idea of infinity pervades mathematical thinking. Infinity makes mathematics possible. Modern mathematics itself did not even begin until mathematicians began to come to terms with the infinite. The first partially successful attempts to deal with infinite sets did not occur until the latter half of the 19th century—surprisingly late. Part of the difficulty in accepting the idea of infinity is that human intuition, honed by a lifetime of familiarity with finite collections (of people, trees, stars, automobiles, and so on), leads us astray when we attempt to formulate even a simple true statement about an infinite collection of mathematical objects. Applied to the infinite, true statements often sound false and false statements often sound true. Mathematical pioneers who bravely tried to understand the concept of the infinite were often ridiculed for their efforts. Today there is a large body of mathematical work on the properties of infinite sets. The ideas that were generated in the attempt to understand the mathematical concept of infinity have had a profound impact on the development of science and technology. These ideas have helped shape the finite world in which we live.

Philosophically inclined individuals in many different places and times have thought about infinity, but the first culture to attempt to study the infinite in a rigorous way was that of the ancient Greeks. They made only modest progress. The Greek philosopher Zeno of Elea (ca. 490 B.C.E.–ca. 435 B.C.E.) used the concept of infinity in a series of four famous logical arguments. Little is known about Zeno except that he was a controversial figure. His ideas made quite an impression on the philosophers of his time, and they have resonated among the mathematically and philosophically inclined ever since. The so-called paradoxes of Zeno were made to bolster a certain philosophical position about the nature of the universe. Zeno, as his teacher Parmenides did,



The rigorous study of infinity began in ancient Greece. (Library of Congress, Prints and Photographs Division)

believed that all of reality is the manifestation of a single being. Because everything is part of a single, eternal, changeless being, change is impossible. Proving the truth of this idea was Zeno's goal, but a positive proof is impossible. Zeno could not prove that he was right, so he did the next best thing: By a clever use of the mathematical concept of infinity he sought to demonstrate that any other worldview leads to a logical contradiction.

Historical sources indicate that Zeno proposed a large number of these paradoxes, but only four have come down to us. These four paradoxes all use some aspect of the infinite to arrive at a conclusion that contradicts our view of reality. One of the arguments, called the Dichotomy, purports to prove the nonexistence of motion. (Motion involves change, and, according to Parmenides, change is impossible.) To appreciate the paradox, imagine a runner just beginning a race. To reach the finish line, the runner must first run half the race and cross the line that is midway between the starting point and the finish line. To reach the midpoint, however, the runner must first run one-fourth of the total distance and cross a point that is halfway between the midpoint of the racecourse and

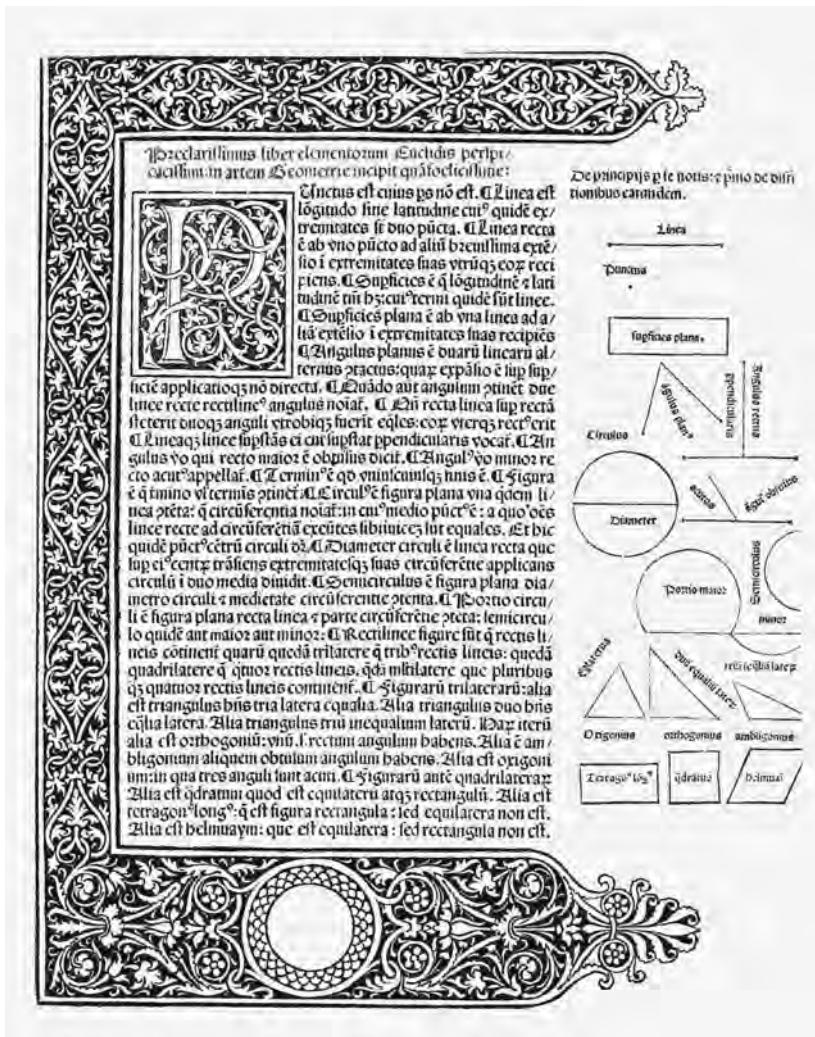
the starting point. This is one-fourth the total distance, but before the runner can run one-fourth of the total distance, the point halfway between a quarter of the distance and the starting point must be crossed. This point is one-eighth the total distance. The pattern goes on and on.

The object of the paradox is to demonstrate that there exist infinitely many “halfway points” between the starting line and the finish line. Furthermore reaching each one of these points must take an “instant” of time. It does not really matter how long an instant is; it is “clear” to most of us that it must take some time to pass each of the points that Zeno has identified. Since it takes an instant to pass each of these points, and since there are infinitely many points, running the race must take an infinitely long time. (An infinite collection of very short time intervals must still add up to an infinitely long time.) The only possible conclusion is that motion is impossible.

The other three paradoxes of Zeno have a similar flavor. Each argument is negative in the sense that it is designed to arrive at a contradiction. Each paradox uses the concept of infinity as a sort of logical sword: Zeno wants to demonstrate that philosophies different from those of Parmenides contain logical fallacies. The only “logical” conclusion, according to Zeno, is to abandon one’s own philosophy and adopt that of his teacher. There is no indication that Zeno’s paradoxes won many converts for Parmenides. Some philosophers simply dismissed them out of hand as meaningless wordplay. The rest were stymied. Refuting Zeno’s paradoxes requires quite a bit of insight into mathematics.

From the point of view of a history of infinity, the Dichotomy is notable in that it fails to convey much information about the nature of infinity. Inherent in the Dichotomy is the assertion that infinitely many instants of time—no matter how small an instant of time is required to pass each point along the race track—add up to an infinitely long time. But what does *infinitely many* mean? How can we distinguish infinite sets from large finite ones? Zeno’s paradoxes are of little help.

Perhaps the most important result about infinite sets to come out of the Greek tradition is to be found in the ninth book of



Title page of Euclid's Elements. Euclid's proof that there exist infinitely many prime numbers was one of the first proofs explicitly involving infinite sets. (Library of Congress, Prints and Photographs Division)

Euclid's *Elements*. This work, a multivolume set of mathematics books, is probably the most-read and most-influential work in the history of mathematics. It was written by the mathematician Euclid of Alexandria. We know little of Euclid, except that he lived

in Alexandria, Egypt, during the third century B.C.E. We know that his book was written to be used as a textbook and that it contains a careful exposition of mathematics as it was known to the Greek mathematicians of his time. In the ninth book of *Elements* Euclid proves that there are infinitely many prime numbers. The proof that Euclid gives of this fact is remarkable in a number of ways. In fact Euclid's proof that there are infinitely many prime numbers is a wonderful snapshot into the way the ancient Greeks understood the concept of an infinite set. Euclid's writings on this subject offer us an opportunity to see how infinity was perceived by the mathematicians of his age.

To understand Euclid's proof that there are infinitely many prime numbers, one needs to be aware of a few facts about the natural number system. The first fact is a definition: A *prime number* is a number that is evenly divisible only by itself and 1. Some examples of prime numbers are 2, 3, 5, and 7. This definition of prime number breaks up the set of all natural numbers greater than 1 into two nonoverlapping classes. Every natural number greater than 1 is either prime or not prime. A number that is not prime is called a composite number. (The number 1 is, technically speaking, neither a prime number nor a composite number. It is called a unit.)

The second important fact is about composite numbers. Every *composite number* is, as a matter of definition, divisible by at least one other number besides itself and 1. What the Greeks had already discovered is that every composite number can be divided by some prime number. There are some composite numbers that are also divisible by composite numbers, but *every* composite number must be divisible by at least one prime. From these observations it is easy to see that the prime numbers occupy a special place in our number system: *Every* natural number either is a prime number or is evenly divisible by a prime number.

The last preparatory fact to keep in mind is that no one has ever found a formula that enables the user to list all prime numbers. For many centuries mathematicians have searched for such a formula, a formula that would make sense out of the set of all prime numbers, but the formula, if it exists, remains elusive. It

is therefore not possible to use a formula that enables one to “produce” all prime numbers.

By way of contrast we can easily state a formula that allows us to list all even numbers. Imagine that we make a numbered list of all the even numbers. There are many ways of doing this; the most direct way is to list them according to size: Write the number 2 in the first place on the list, the number 4 in the second place on the list, the number 6 in the third place on the list, and so on. In fact we can generalize this scheme and simply say that we write the number $2n$ in the n th place on the list, where we let n represent any natural number—“the number $2n$ in the n th place” is our formula for listing all even numbers. At present if we want to list all primes in the same way that we list all even numbers, we can move down a list of all natural numbers checking each number as it appears to see whether there are other numbers besides itself and 1 that evenly divide it. Initially this is easy because the numbers involved are small, but as the numbers become larger and the list of primes becomes longer, the amount of computation involved in determining whether or not a number is prime becomes substantial. The method we used to show that there are as many even positive integers as there are natural numbers is not suitable for proving that there are infinitely many primes.

This is the condition Euclid faced when he wanted to know whether the set of all primes contains infinitely many numbers. Euclid’s approach to this problem is very telling. He does not try to prove directly that the set of all prime numbers is infinitely large. Instead he uses an indirect approach. He begins by *assuming* that the set of all prime numbers contains only finitely many primes. This is his crucial assumption, because *if* there are only finitely many primes and he makes a list of all the prime numbers, then his list, at some point, ends. No matter how he lists the primes—according to size or some other criterion—his assumption that there are only finitely many primes assures that the list must have a last element.

Euclid imagines making such a list. It would have to look something like this: $p_1, p_2, p_3, \dots, p_n$, where each letter represents

one of the prime numbers in the set and p_n represents the last of the primes on the list. (The subscripts refer to the position on the list where each prime appears.) Presumably there are many primes, so the number n , which indicates how many numbers are in the set, is a very large number. The size of n is, however, less important than the fact that it exists at all. Euclid's next step is to "build" a new number. We can call this number M . We obtain M by multiplying all the prime numbers on our list together and then adding 1 to the result: $M = p_1 \times p_2 \times p_3 \times \dots \times p_n + 1$. This is the proof. All that remains is to figure out what Euclid has done.

Notice that M is bigger than any number on the list of primes, so it cannot be a prime number itself, since by assumption Euclid's list contains *all* prime numbers. The difficulty that arises is that M also cannot be a composite number. To see why, recall that *every* composite number is divisible by a prime number, so if M is composite then it must be divisible by one of those prime numbers. But it is not. If we try to divide M by p_1 , for example, our quotient is $p_2 \times p_3 \times p_3 \times \dots \times p_n$ plus a remainder of 1. This shows that the prime p_1 does not divide M evenly. In fact no matter which prime on our list we choose to divide into M , the result is always the same. We always get a remainder of 1 in our answer. Since every number greater than 1 is either prime or composite—and Euclid has proved that M is neither prime nor composite—he has obtained a contradiction. Because the conclusion is false, the premise must also be false. The premise—that there exist only finitely many primes—must be rejected. Euclid concludes that there are infinitely many primes.

This is a remarkable proof. It is remarkable because it tells us virtually nothing about infinite sets. The proof is constructed in such a way that all one need know about infinite sets is that they are not finite. To be sure, that is not very much insight into infinite sets, but that is all that Euclid needs. He shows only that when he assumes that the collection of all prime numbers is finite he obtains a logical contradiction. Since every set is either finite or infinite, the only other possible resolution for the contradiction is that the set of prime numbers is infinite. Euclid's proof is a good

illustration of what the Greeks understood about infinite sets: They knew that infinite sets are not finite, but they did not know much else about them.

Euclid's proof also is one of the very first instances of proving a mathematical statement by contradiction. This is a common logical technique today. The mathematician assumes something that is false and then shows that the assumption necessarily leads to a false conclusion. If the logic between the assumption and the conclusion is without error then the only explanation is that the assumption made at the outset of the proof is false. Because the assumption is false one can safely conclude that the opposite of the assumption must be true.

The Greeks, who were never comfortable with infinite sets, produced a tremendous amount of mathematics. Since infinity is a concept that is central to any number system, the question arises, How did they manage to do mathematics without the idea of an infinite set? The answer lies in the observation that there are always as many numbers as we need. It is apparent to anyone who knows arithmetic that any number different from 0, no matter how small, can always be diminished in size by multiplying it by 1/2. Similarly any number, no matter how large, can always be increased by adding 1.

These procedures can be continued indefinitely. We never reach a largest number past which we cannot go. Nor does the procedure of multiplying a number different from 0 by 1/2 ever yield the number 0. In this sense no one who deals with numbers can escape the idea that the number system is infinite. The Greeks, however, were careful to distinguish between infinite sets and what we might call infinite processes. An *infinite process* is a procedure consisting of a series of steps. We can repeat the procedure as often as we like. There is no theoretical limit, aside from the finiteness of life, to repeating the procedure any number of times. We have just encountered two examples of such processes: (1) the procedure of adding 1 at the n th step and (2) the procedure of multiplying by 1/2 at the n th step, where n can represent any natural number. In this view of infinity, the process is never completed. Although we do not need to stop at any particular step, we

must eventually stop at some step. Aristotle expressed the idea in this way:

Our account of the infinite does not rob the mathematicians of their science, by disproving the actual existence of the infinite in the direction of increase. . . . In point of fact they do not need the infinite and do not use it. They postulate only that the finite straight line may be produced as far as they wish.

(Aristotle. Physica. Translated by R. P. Hardie and R. K. Haye. Oxford: The Clarendon Press, 1930. Reprinted by permission of Oxford University Press.)

By *produced* Aristotle means extended.

A classical example of the idea of an infinite process is the method of exhaustion, which was first proposed by the Greek philosopher Eudoxus of Cnidus (ca. 408 B.C.E.–ca. 355 B.C.E.). To understand the method of exhaustion, imagine that we begin with some large number. Our goal is to reduce the size of this quantity in steps. Our first step is to multiply our large number by a number no greater than $1/2$. Our answer is the “output” from step one. At each successive step we multiply the output obtained at the previous step by a number no greater than $1/2$ to obtain the output for the present step. If we do this often enough we are left with a quantity that is “small.” Of course, this has little meaning until we make clear what we mean by the word *small*.

To make these ideas precise, let us choose some positive number to represent our idea of small. We might choose the number 0.1 or the number 0.000001. Now we simply reduce the size of the original large number such that after each step the remaining number is no more than half the size it was at the preceding step. After a finite number of steps, we have reduced the size of our original number to something that is less than the number we use to define *small*. At this point we can continue to reduce the given number for as long as we want, but once our number is small enough—that is, once it is smaller than the number we choose to describe our idea of what *small* means—we have reached what computer programmers define as an *exit condition*: When our

remainder is small enough, we simply stop. This concept of an infinite process is one that the Greeks found useful, and it enabled them to avoid the manipulation of sets that contained infinitely many elements.

The Greek preference for what is sometimes called the “potentially infinite” over the “actually infinite” was shared by many generations of mathematicians. Islamic mathematicians adhered to this idea, and European mathematicians used this approach well into the 19th century. The manipulation of infinite sets is logically tricky. The properties of infinite sets are often nonobvious, even surprising. Hence for thousands of years after the Greeks mathematicians handled the problem of infinity largely by avoiding it.

10

GALILEO AND BOLZANO

The first important breakthrough in understanding the nature of infinite sets was made by the Italian scientist, inventor, and mathematician Galileo Galilei (1563–1642). Galileo was the son of the musician and composer Vincenzo Galilei, whose compositions are still occasionally performed and recorded. Growing up, Galileo had very little exposure to mathematics. When he enrolled in the University of Pisa at age 17, he intended to study medicine. While at the university he is said to have overheard a geometry lesson, and this chance experience marks the beginning of his study of mathematics and science.

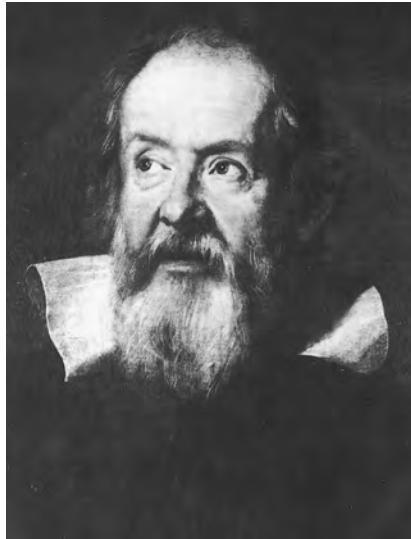
Before he could graduate from the University of Pisa, Galileo ran out of money and had to withdraw. Whatever he had learned at the university, however, was enough to get him started as a scientist and inventor. Soon he was famous as both, and within four years of dropping out of the University of Pisa he was back—this time as a member of the faculty.

Today Galileo is remembered for several reasons. First, his astronomical observations of the Moon, Venus, and Jupiter changed the way the people of his time perceived their place in the universe. For philosophical and religious reasons most of the people of Galileo's time and place believed that Earth is the center of the universe. Galileo did much to discredit these ideas. Discrediting the teachings of the Catholic Church was, at this time, a very dangerous undertaking. Galileo was threatened with torture because of his ideas. He spent the last several years of his life under house arrest. Galileo is remembered for the grace and courage with which he faced his persecutors as well as for his ideas.

Galileo was also one of the first to combine higher mathematics with experimental physics in a way that is truly modern. His ideas on these matters are contained in the book *Dialogues Concerning Two New Sciences*, which was written near the end of his life. (During this time he was under house arrest, and as part of his punishment he had been ordered to stop writing about science.) In this book Galileo also wrote about the nature of infinite sets.

The observations that Galileo made about infinite sets are fairly brief. He enjoyed mathematics, but he was primarily a user of mathematics—Galileo knew how to employ the mathematics that was known at the time to help him see further into scientific problems. Today we might call him an applied mathematician, because, for the most part, Galileo used mathematics as a tool. Fortunately he also had an eye for interesting mathematical relationships and mathematical ideas. When he spotted an interesting mathematical fact or relationship he did not hesitate to point it out to others, but he did not spend much of his time investigating mathematics as a distinct branch of knowledge. He was, at heart, a scientist.

Galileo's mathematical descriptions in *Dialogues Concerning Two New Sciences* are very wordy by modern standards because his book is a work of fiction. It is a literary work as well as a scientific one, written in the form of a conversation among three characters. The characters, each of whom has a distinct personality, spend four



Galileo Galilei was the first to notice what would eventually be regarded as the defining property of infinite sets: An infinite set can be placed in one-to-one correspondence with a proper subset of itself. (Library of Congress, Prints and Photographs Division)

days discussing the ideas on which the new sciences are founded. Each chapter comprises one day in the book. As the title implies, most of the conversation is scientific, but there are nonscientific digressions as well. Some of the dialogue concerns philosophical matters, and some of it is of a mathematical nature.

At one point the characters discuss infinite sets. Through this discussion Galileo describes a very interesting and important property of infinite sets. He seems to have been the first person in history to notice this extraordinary property—at least he was the first person to write about it—a property that plays a very important role in the identification and use of infinite sets. Salviati, the scientist, who is evidently a stand-in for Galileo himself, explains his ideas about the infinite to the ever-practical Sagredo and their dim-witted friend Simplicio. Salviati makes the discussion as concrete as possible by considering a numerical example. His goal is to prove to his friends that there are as many perfect squares as there are natural numbers. (A *perfect square* is a number whose square root is a natural number. The numbers 4 and 9, for example, are perfect squares, because their square roots are 2 and 3, respectively. The number 5, on the other hand, is not a perfect square, because $\sqrt{5} = 2.236 \dots$.)

The assertion that there are as many perfect squares as natural numbers requires proof, because it violates commonsense notions. Remember that all perfect squares are themselves natural numbers, but many natural numbers are *not* perfect squares. Mathematically we describe this situation by saying that the set of all perfect squares is a *proper subset* of the set of natural numbers. Galileo's assertion that there are as many perfect squares as natural numbers means that if we remove the set of numbers that are *not* perfect squares from the set of all natural numbers, the set of numbers remaining is the same size as the original set. The only possible conclusion is that for the problem under consideration—and consequently for infinite sets in general—sometimes the whole set is no larger than a part.

From the dialogue of the book, it is clear that Galileo is fascinated by this observation. The main character, Salviati, even goes to some lengths to demonstrate to his friends how few numbers

are, in fact, perfect squares. He points out that only a 10th of the numbers between 1 and 100 are perfect squares and only 1 of 1,000 numbers between 1 and 1 million is a perfect square. Generally the larger the interval he considers, the smaller the percentage of numbers that are perfect squares. So, the simple Simplicio wonders, how can it be that there are as many perfect squares as integers?

In the end Galileo's method of proof reduces to making a numbered list. The first position on the list is occupied by the first perfect square, the number 1. The next perfect square on the list—the perfect square in second position—is 4. Next up in third position is the number 9. The general rule is that if we let the letter n represent any natural number, then the perfect square in the n th position on the list is the number n^2 . Because Salvati's list of perfect squares never ends—every position on the list is occupied by exactly one perfect square—he concludes that there must be exactly as many perfect squares as there are integers. This is an imaginative and bold conclusion, because it requires us to accept a property that is peculiar to infinite sets, a property that violates our ordinary, everyday experience.

For an infinite set it is often the case that a part of the set can be placed in one-to-one correspondence with the whole in the following sense: We can pair up a subset with the parent set in such a way that every element in the parent set is paired up with a unique element in the subset and no members of either set are left over. In Galileo's example the "part" consists of the set of perfect squares. The "whole" or parent set consists of the entire set of natural numbers.

1.)	1
2.)	4
3.)	9
4.)	16
5.)	25

⋮ ⋮

⋮ ⋮

n.)	n^2
-----	-------

⋮ ⋮

⋮ ⋮

Galileo observed that there are as many perfect squares as natural numbers although infinitely many natural numbers are not perfect squares.

This is an important observation about infinite sets, and it is well worth our while to explore the idea a little. First, given any infinite set, we can always remove any finite collection of elements without changing the size of the original, parent set. For example, suppose that we remove the first 10 natural numbers from the set of all natural numbers. These 10 deletions leave the size of the set unchanged. To prove this is so, we simply follow Galileo's example and make a numbered list. In first position we write 11, in second position we write 12, in third position we write 13, and in n th position we write $10 + n$. Since every position on the numbered list is occupied by an element from our set we have proved that the set of all natural numbers greater than 10 is just as large as the set of all natural numbers. Second, it is sometimes even possible, as Galileo himself demonstrated, to remove infinitely many elements from an infinite set without changing the size of the set, although this must be done with some care. The property that the whole can equal the part is true only for infinite sets. By contrast for finite sets it is never true that a proper subset is the same size as the parent set from which it is drawn.

Galileo's method for comparing the set of natural numbers with the set of perfect squares also bears mention. Normally when we want to compare the size of two finite sets we count them, but counting is an ineffective tool for investigating infinite sets: Infinite sets are too large. We can try counting the set of natural numbers, but no matter how many we count and no matter when we stop, most natural numbers remain uncounted. It is to Galileo's credit that he recognized the importance of establishing a correspondence between a set whose properties he more or less understood, the set of natural numbers, and a set with whose properties he was less familiar, the set of perfect squares.

The central idea of correspondence is easily illustrated if we imagine a movie theater just before the lights are dimmed. It is a simple matter to check for empty seats. If we see empty seats and there is no one left standing then we know that the set of seats is larger than the set of people in the theater. If, however, there are no empty seats and there are a few people standing in the back, then we can be sure that the number of people exceeds the

number of seats. If all of the seats are occupied and no one is left standing then there is a one-to-one correspondence between the number of people and the number of seats. In this case we can be sure that there are exactly as many people as seats. All of our conclusions about the relative numbers of people and seats are valid *even though we may not know either the number of seats in the theater or the number of people.* This is Galileo's insight. It allows him to compare the relative sizes of sets without knowing the true, or "absolute," size of either one. He has no real insight into how large the collection of natural numbers is, but by establishing a correspondence between the set of natural numbers and the set of perfect squares by his listlike approach, he is able to prove that both sets are of equal size.

Galileo's insights are very important to the theory of infinite sets, but Galileo himself does not know what to do next. Having discovered that a proper subset of an infinite set may be exactly the same size as the parent set, Galileo hits a conceptual dead end. He clearly recognizes that he has made an important discovery, but he is unsure about what conclusions can be drawn from it. In the end he concludes that if a proper subset of an infinite set is the same size as the set from which it is drawn, then the ideas of "greater than," "equal," and "less than" have no place in a discussion of the infinite. This is how the character Salviati describes the situation:

So far as I see, we can only infer that the number of squares is infinite and the number of their roots is infinite; neither is the number of squares less than the totality of all numbers, nor the latter greater than the former; and finally the attributes "equal," "greater" and "less" are not applicable to infinite, but only to finite quantities.

(Galileo Galilei. Dialogues Concerning Two New Sciences. Translated by Henry Crew and Alfonso de Salvio. New York: Dover Publications, 1954, page 88.)

Galileo's arguments are all correct. Only his conclusion is wrong. We soon see that "greater," "less," and "equal" do, indeed, belong to the study of the infinite.

INFINITY AS A NUMBER

The first serious attempt to work out a mathematics of infinity was made by the mathematicians of the Indian subcontinent. The Indian mathematician and astronomer Bhaskara II (1114–ca. 1185) describes the thinking behind this arithmetic of infinity. (The roman numeral II in Bhaskara's name is used to differentiate him from another prominent mathematician of the same name. We will make no further reference to Bhaskara I.) Bhaskara was one of the preeminent mathematicians in the Hindu tradition. In addition to working as a mathematician he was head of the astronomical observatory at Ujjain, the leading Indian observatory of the time. His writings contain most of what is best in classical Indian mathematics. He demonstrates a thorough knowledge of the algebra that was of interest to the mathematicians of his time as well as an appreciation and mastery of the place-value system of notation, the innovation for which these mathematicians are best remembered today.

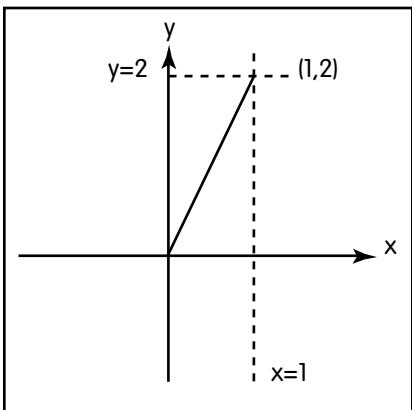
Bhaskara is thoroughly familiar with the rules of ordinary arithmetic. He knows how to work with fractions and signed numbers. He says, for example, that the product of two numbers of the same sign is positive and the product of two numbers of opposite sign is negative. He is also somewhat familiar with arithmetic operations involving 0. These ideas had long been part of Indian mathematics, but Bhaskara wants to go further. He attempts to introduce the concept of infinity as a way of "completing" arithmetic. To understand his thinking, let x represent a positive number. Consider the fraction $2/x$. The smaller the chosen value of x is, the larger the fraction $2/x$ becomes. In fact we can make the quotient $2/x$ as large as we want provided we make x small enough. Furthermore for any positive number x , we can always

Galileo's writings on infinite sets had little influence on his contemporaries. For the next few centuries mathematicians continued to shun infinite sets. The next mathematician to make a serious study of the infinite was a little-known Czech priest and mathematician, Bernhard Bolzano (1781–1848). Bolzano graduated from the University of Prague in 1805, when he was ordained a priest and received a Ph.D. in mathematics. He then accepted a position as professor of philosophy and theology at the university.

"recover" the numerator by multiplying the quotient by x : $x \times 2/x = 2$. Bhaskara wanted to extend these ideas to the case in which x represents the number 0. He begins by asserting that $2/0$ is infinity. (In modern notation the idea of infinity is often represented by the symbol ∞ ; Bhaskara asserts that $2/0 = \infty$.) The reason for this assertion is that as long as x remains greater than 0, $2/x$ represents an ordinary number. The closer x is to 0, however, the larger this quotient becomes. Bhaskara's idea is that as x approaches 0, $2/x$ also approaches a number and he calls this number ∞ .

Bhaskara indicates that we can recover the 2 when x equals 0 in exactly the same way that we recover 2 when x is not 0. He asserts that when x equals 0 we need only multiply the quotient by 0 and we get the equation $0 \times 2/0 = 2$, or, to put it another way, $0 \times \infty = 2$. If this equation were true then we could do arithmetic with the number 0 in just the same way that we do arithmetic with every other number, but Bhaskara's reasoning is flawed. The problem is that we can repeat every word of this paragraph and simply substitute, for example, the number 3 for 2, and if what Bhaskara says about 2 is true, then what we say about 3 must also be true. In fact any positive number greater than 0 can be substituted for 2, and if the previous statements are true of 2, then they must be true of every other positive number. Let a represent any positive number. According to Bhaskara $a/0 = \infty$, but then, when we try to recover a by multiplying both sides of the equation by 0 we arrive at the result that $0 \times \infty$ could be equal to any positive number. As a consequence $0 \times \infty$ has no meaning. Trying to divide by 0 leads to contradictions, and this is the reason Bhaskara's ideas were rejected. Today mathematicians reject the possibility of dividing by 0. One cannot simply "complete" the number system in the way that Bhaskara envisions. Infinity cannot be treated as if it were an ordinary, finite number.

He was a determined and highly creative man. In an age of militarism Bolzano advocated nonviolence and wrote about the futility of war. His ideas were very unpopular with those in authority, both secular and religious, but Bolzano persisted in writing about the necessity of reforming and demilitarizing Czech society despite pressure to desist. In 1819 Bolzano was fired from his position at the University of Prague. He even spent some time under house arrest.



Notice that every x in the interval $[0, 1]$ is paired with exactly one y in the interval $[0, 2]$, and that every y in the interval $[0, 2]$ is paired with exactly one x in the interval $[0, 1]$. This shows that there are as many points in the shorter interval as there are in the longer one.

The loss of his position at the university freed Bolzano to devote himself to philosophy and mathematics. His mathematical goals were very ambitious. He wanted to establish the field of mathematics on a more rigorous foundation. In this regard he was very forward-thinking, and today he is generally regarded as a very far-sighted mathematician. During his lifetime, however, his ideas received little recognition.

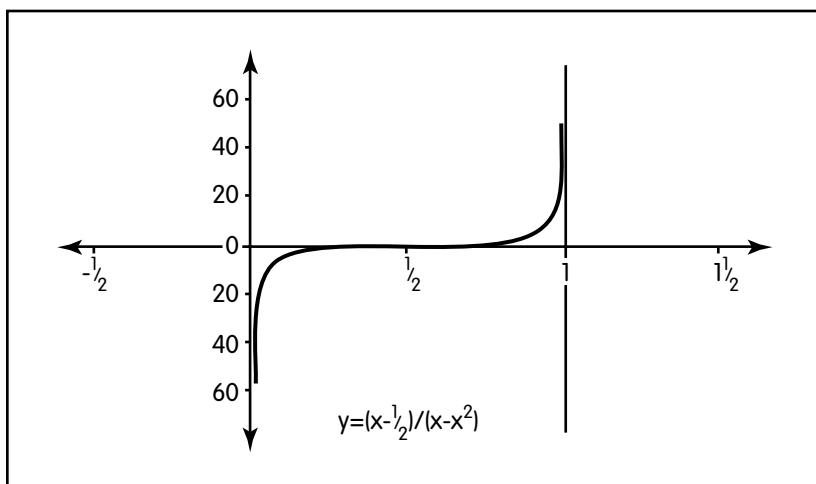
Bolzano recognized the importance of infinite sets and spent some of his energy studying their properties. Like Galileo he was fascinated by the fact that a set can be

put into one-to-one correspondence with proper subsets of itself. Unlike Galileo, who wrote about the natural numbers, Bolzano was interested in what we call the set of real numbers. Mathematicians of his time still had only a hazy idea of what a real number is. We can understand Bolzano's ideas about infinite sets of real numbers by using simple graphs.

Consider the interval on the real line consisting of all numbers that are greater than or equal to 0 and less than or equal to 1. This interval is one unit long and contains the points 0 and 1. Next draw a sloping line above this interval. Notice that each point on the line has an x -coordinate no smaller than 0 and no greater than 1. How much the y -coordinates of the points on the line vary depends on the slope of the line. In our picture the y -coordinates begin at 0 and end at 2. This graph shows that we can place all of the points between 0 and 1 into one-to-one correspondence with the set of all points between 0 and 2. This shows that although the

segment containing the y -coordinates is twice as long as the segment containing the x -coordinates, there are as many points on the shorter segment as there are on the longer one. In fact we can make our line slope as steeply as we want. The result is a one-to-one correspondence between the set of points in the interval with endpoints 0 and 1 and the interval with endpoints 0 and n , where we can let n represent any natural number greater than 0.

We can even go further. Consider the equation $y = (x - 1/2)/(x - x^2)$ between the numbers $x = 0$ and $x = 1$. (This function is not defined when x is 0 or when x is 1 because at those points the denominator is 0.) The graph of this function is shown in the accompanying diagram. It extends upward and downward along the entire length of the y -axis. Here we have a one-to-one correspondence between the set of real numbers between 0 and 1 (exclusive) and the entire y -axis. This shows that there are as many points between 0 and 1 as there are on the entire real line. Nor is there anything special about the interval between 0 and 1. A



An illustration of Bolzano's observation that any interval of the real number line can be placed in one-to-one correspondence with the numbers along the entire line: Every point on the interval (0, 1) has been paired with exactly one number on the y-axis and every number on the y-axis has been paired with exactly one number from the interval (0, 1).

THE LIFE AND OPINIONS OF TRISTRAM SHANDY, GENTLEMAN

Sometimes art is better at conveying the concept of the infinite than is mathematics. Although the word *infinite* is familiar to all, the concept is so far beyond our ordinary experience that our intuition is often a poor guide in developing insight into what mathematical infinity means. The mathematician, philosopher, and logician Bertrand Russell illustrated the difference between the finite and the infinite with the tale of Tristram Shandy, and we follow his practice here.

The Life and Opinions of Tristram Shandy, Gentleman, is a novel by the Irish-born English novelist, humorist, and Anglican priest Laurence Sterne (1713–68). *Tristram Shandy* was a radical approach to storytelling for its time. The punctuation, the organization of the story, the stream-of-consciousness style, and the subject matter—the story is the first novel about writing a novel ever written—make the story a remarkable achievement. It was published in nine volumes over a period of eight years (1759–67) and proved to be very popular. The narrator in the novel, Tristram Shandy, has decided to tell the story of his life beginning at the time of his conception. Tristram has a keen eye for detail, and he frequently interrupts his narrative to relate a story about his quirky family. In fact he has so much to relate to

similar sort of argument shows that there is a one-to-one correspondence between the points contained in any interval, no matter how small, and the entire real axis. These types of paradoxes made many mathematicians suspicious of the entire subject.

Bolzano coined the word *set*, and this simple act is also important for the subject. First, with the concept of set, mathematicians had a convenient way of imagining the entire collection of natural numbers. They were all elements in a single set. More important, Bolzano described a *set* as a collection of objects whose definition does not depend on the order in which its elements appear. This point emphasizes how fundamental the concept of correspondence is. There are two ways of comparing the sizes of finite sets: One way is to develop a correspondence between the elements of the two sets; the second way is to count them. Counting

the reader that he does not get around to describing his birth until volume four of the series. Eventually, however (in volume 9), Shandy realizes that his situation is hopeless: It is taking longer to tell the story of his life than he has time to live it, and the novel ends.

Mathematically speaking the problem Shandy faces is not that he is taking so long to tell the story of his life; it is that he has so little time to tell it. If Shandy had infinitely many years to tell his story, then despite the fact that he continued to fall further and further behind, he would, in the end, have sufficient time to complete the story. It may seem that living longer would only make the situation worse, but this is not the case. Even if it took 1,000 years to relate one year of his life, he would still have enough time to tell the story if he lived forever. After 2,000 years he would have completed the first two years of his life, after 3,000 years he would have completed the third year, and so on. In this manner we can establish a correspondence between each thousand years of description and the one year of life he purports to describe. The idea is identical to Galileo's correspondence between perfect squares and natural numbers. Our correspondence would look like this: (1, 1000), (2, 2000), (3, 3000), . . . What is important is that each year of his life would appear somewhere on this list. The fact that he continues to fall further and further behind is irrelevant if he has infinitely many years remaining. The longer Tristram lives the more he can tell. If he could live long enough, he could tell us everything.

a set has significant drawbacks. One disadvantage of counting is that it requires us to introduce an order onto the elements of the set. When we count a finite set we must choose a first element and a last. (In the case of an infinite set there is no last element.) Order is an additional restriction that we impose from without. It has nothing to do with the size of the set. Establishing a correspondence allows us to compare sets without introducing an order. Unfortunately none of these important ideas had much influence on Bolzano's contemporaries.

The investigations of Galileo and Bolzano into infinite sets were similar in concept. We soon see that the principal difference between the two is that Bolzano investigates sets that are larger than those of Galileo. Both individuals, however, recognize the importance of the idea of correspondence, and both recognize that



A scene from The Life and Opinions of Tristram Shandy, Gentleman, a story that is sometimes used to demonstrate how ideas about the infinite can conflict with our commonsense notions. (Library of Congress, Prints and Photographs Division)

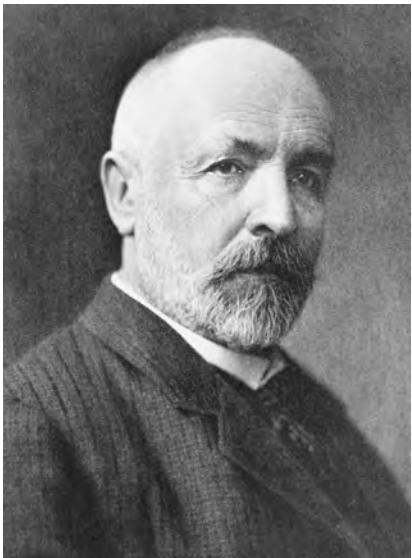
in the case of infinite sets it is entirely possible that a proper subset can be the same size as the parent set from which it is drawn. Finally, neither Galileo nor Bolzano goes much beyond these observations. It would be a few more decades before anyone understood how to found a theory of the infinite based on the observations made by these two pioneers.

11

GEORG CANTOR AND THE LOGIC OF THE INFINITE

The great breakthrough in the theory of sets and in the history of mathematics in general occurred with the work of the German mathematician Georg Cantor (1845–1918). Cantor's insights into the infinite and their importance to the history of mathematics were not immediately appreciated. During much of his life his efforts were ridiculed by many other prominent mathematicians, and he was blocked from a position that he very much wanted at the University of Berlin by one of his former teachers, who was opposed to Cantor's attempts to understand and use infinite sets. Still Georg Cantor did not give up. Today Cantor's work is often described as forming the foundation of the modern age of mathematics.

Cantor was born in Saint Petersburg, Russia, of Danish parents. His family was prosperous, and they remained in Saint Petersburg until 1856, when his father became ill, and they moved to the German city of Frankfurt. During his teen years Cantor showed unusual interest and ability in mathematics. He received most of his education at the University of Berlin, where he was taught by some of the finest mathematicians of the day, including Karl Theodore Weierstrass and Leopold Kronecker. Cantor's doctoral thesis has the engaging name “In Re Mathematica Ars Propendi Pluris Facienda Est Quam Solvendi” (In mathematics the art of asking questions is more valuable than solving problems). Soon after graduation Cantor joined the faculty of the University of Halle, in Wittenberg, Germany, where he remained for his entire working life.



Georg Cantor; whose discoveries about the nature of infinite sets had a profound influence on the development of mathematics. (CORBIS)

Cantor was not immediately drawn to the theory of infinite sets and their properties. He became interested in the mathematical concept of infinity after discussions with Richard Dedekind, the mathematician who developed a sound logical basis for the real numbers. Perhaps their discussions were entirely mathematical, but interest in mathematics was not his only motivation for his study of the infinite. Cantor's interest in the infinite was also partly theological. Religion must have been an important topic in Cantor's boyhood home. His father was Protestant and his mother was Catholic at a

time when this type of "mixed marriage" was both unusual and controversial. Cantor seems to have believed that insights into the infinite are also insights into God's handiwork. In any case Cantor and Dedekind became lifelong friends, and, when it was possible, Dedekind offered support to Cantor and his controversial ideas.

Cantor's first goal was to identify that aspect that every infinite set shares and that distinguishes all infinite sets from all finite ones. In his proof that the set of all prime numbers is infinite, Euclid simply states that infinite sets are not finite. That is a good enough criterion for the purpose of completing a single proof, but it is not possible to build a comprehensive theory of anything, including infinite sets, based on what it is not.

To begin his theory of infinite sets, Cantor makes a bold decision. He begins his theory by rejecting one of Euclid's most fundamental axioms: "The whole is greater than the part." This

simple sentence clearly conforms to our everyday experience. If we spend or consume or tear down part of what we have, whatever remains is less than that with which we begin. Before Cantor practically all mathematicians and nonmathematicians alike accepted this idea as a truism. Rejecting Euclid's axiom is necessary, however, because if we use the axiom in the study of the infinite, we cannot prevent logical contradictions. In fact it is precisely the converse of this statement, namely, "There are parts that are equal to the whole," that serves to distinguish all infinite sets from all finite ones. If we accept that for infinite sets there are parts that are equal to the whole, then, instead of encountering logically contradictory consequences, we encounter only strange consequences, and strangeness is no barrier to good mathematics. It is only logically contradictory ideas that must be avoided. Cantor defines infinite sets as those sets that have proper subsets that are the same size as the parent set. This definition originates with Dedekind, but Cantor makes greater use of it. The consequences of this definition are disconcerting enough that Cantor's ideas alienated many mathematicians for years.

To state Cantor's definition of an infinite set more formally, we can say that a set is *infinite* provided it can be placed in one-to-one correspondence with a proper subset of itself. Further we call two sets, infinite or not, the same size provided their elements can be placed in one-to-one correspondence with each other. Comparing sets then becomes a matter of finding correspondences between them.

One of Cantor's early discoveries is that the set of all rational numbers is the same size—or, as mathematicians say, "the same cardinality"—as the set of natural numbers. (Two sets have the same *cardinality* if they can be placed in one-to-one correspondence with each other.) That the rational numbers are no more numerous than the natural numbers must have been something of a surprise. If we imagine the natural numbers and the rational numbers as points on the real line, it is easy to see that between any two natural numbers there are infinitely many rational numbers. In fact between any two distinct points on the real line *no matter how closely together they are spaced* there are infinitely many

rational numbers. It appears, therefore, that there ought to be more rational numbers than natural numbers, but there are not. The proof is not difficult (see the sidebar), but other, more subtle results follow. Cantor soon discovers that the set of real algebraic numbers—that is, the set of all real numbers that are roots of polynomial equations with rational number coefficients—is also the same size as the set of natural numbers. This is even more

THERE ARE NO MORE RATIONAL NUMBERS THAN NATURAL NUMBERS

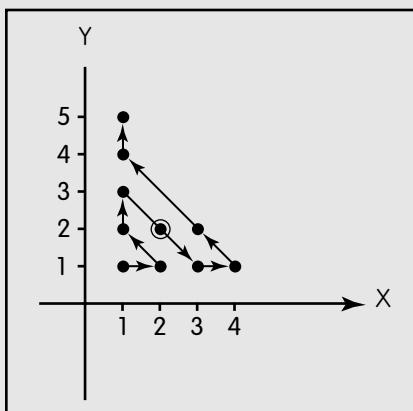


Diagram accompanying the proof that there exist only as many rational numbers as natural numbers.

The fact that there are only as many rational numbers as there are natural numbers surprises many people even today, but the proof of this statement is neither long nor hard. To demonstrate how this is done, we show that there are no more positive rational numbers than there are natural numbers. (The proof is easily extended to show that there are no more positive and negative rational numbers than there are natural numbers.) To establish the correspondence we let a/b represent some rational number, where a and b are natural numbers. Of course there is more

than one way to write a rational number. For example, $1/2$ and $2/4$ represent the same number, so we always assume that our rational number is in lowest terms. To each positive rational number there corresponds a point on the plane of the form (a, b) , where a and b are positive whole numbers, and a/b is in lowest terms, and to each point (a, b) , where a and b are positive whole numbers, there corresponds a unique rational number, c/d , such that $a/b = c/d$ and c/d is in lowest terms. (In what follows, we have to be careful to avoid fractions that are not in lowest terms.)

surprising, because the algebraic numbers encompass all the rational numbers and many irrational numbers as well, for example, $\sqrt{2}$ and $\sqrt[3]{5} + \sqrt{7}$. In fact the set of all algebraic numbers encompasses almost all the numbers with which we are familiar. The two major exceptions are π and e , which are transcendental. (The letter e represents the irrational number 2.71828 . . . , a number that has a very important role in calculus.)

To obtain this correspondence, imagine taking a path through the upper right quadrant of the plane. Our path takes us through each point with positive whole number coordinates. These are called the lattice points of the plane. We use some of these lattice points to form a numbered list. If a point represents a fraction in lowest terms—the point (1, 2), for example, represents the fraction $1/2$ —then we add it to our list. If the point does not represent a fraction in lowest terms—for example, the point (3, 6) does not represent a fraction in lowest terms—then we do not add it to our list. Those points that are added are added in the order in which we encounter them as we travel along our path.

Here is the correspondence: Beginning at the point (1, 1) in the accompanying illustration, follow the path indicated by the arrows. The first point we encounter is (1, 1), and we add it to our list. The second point is (2, 1); this, too, is added to the list. Similarly, the third lattice point, (1, 2), is added to the list. The fourth point, (1, 3), is likewise added to the list, but the fifth point, (2, 2), is not added to the list because it represents the same rational number represented by the point (1, 1). The difference is that the point (2, 2) is not in lowest terms. We continue to list those lattice points that represent rational numbers that are in lowest terms, and we continue to ignore those lattice points that do not represent rational numbers in lowest terms. By following this path we establish a one-to-one correspondence between the natural numbers, or entries on our list, and the set of all positive rational numbers. Every positive rational number is represented by a lattice point, and given any lattice point, our path eventually passes through that point. If the point represents the rational number in lowest terms, we add it to the list; otherwise, we ignore it. This correspondence is the proof that there are only as many positive rational numbers as natural numbers. This famous result is eventually encountered by all serious students of mathematics.

THERE ARE MORE REAL NUMBERS THAN NATURAL NUMBERS

The following proof that the set of real numbers between 0 and 1 cannot be placed in one-to-one correspondence with the set of natural numbers is very similar to the one originally proposed by Cantor. Because there are at least as many real numbers as there are real numbers between 0 and 1, this proof also shows that the set of real numbers is larger than the set of natural numbers.

The first step in the proof is to express every real number between 0 and 1 in its base 10 decimal expansion. The number $1/3$, for example, is written as 0.333 . . . and the number $\sqrt{1/2}$ is written as 0.7071 In fact, any number that has no non-0 digits to the left of the decimal point and at least one non-0 digit to the right of the decimal point represents some real number between 0 and 1. (There is one exception: the number 0.999 . . . consisting of all 9s. This is just another way of writing the number 1, but this makes no difference to our proof.) Next we assume that this set of real numbers can be placed in one-to-one correspondence with the set of natural numbers. If the assumption is true, it is possible to make a numbered list, and next to each natural number there is a real number. Moreover every real number appears somewhere on our list. The list might look something like this:

1. 0.333 . . .
2. 0.7071 . . .

Given the discovery that the set of rational numbers and the set of natural numbers are the same size, it may seem that Galileo is correct in his assertion that “equal,” “greater,” and “less” are not applicable to infinite, or, at the very least, that “greater” and “less” are not applicable to infinite sets. But Galileo is wrong. Cantor shows that the set of real numbers is a larger set than the set of all natural numbers. Like Euclid’s proof that the set of all prime numbers is infinite, Cantor’s proof that the set of all real numbers is larger than the set of all natural numbers begins by assuming the converse, namely, that there exists some one-to-one correspondence between the set of all natural numbers and the set of all real

- 3.** 0.10101 . . .
- 4.** 0.66666 . . .

and so forth. (Of course, there is no need to begin with 0.333 . . . We do so only to provide a concrete example. The following proof shows that no matter how we order the set of real numbers at least one real number fails to make it onto the list.)

To complete the proof we simply construct a real number that is larger than 0 and less than 1 and is not on this list. We construct our number by using only the digits 5 and 6. We begin our number with a 6, namely, 0.6. This ensures that our number is not equal to the first number on the list, which is 0.333 . . . The next digit is also a 6. This ensures that our number is different from the second number on the list in the second decimal place. At this point our number is 0.66. At the third step we add another 6; this time in the 1,000ths place so that we have 0.666. This third step ensures that our number differs from the third number on the list in the third place. In the fourth place we write a 5. This guarantees that our constructed number differs from the number in the fourth place on our list in the fourth decimal position. The process continues indefinitely. In the n th decimal place we write a 6 provided that the n th decimal digit of the n th number is not 6. If the n th decimal digit is 6, then we write a 5.

The number constructed in this manner does not appear anywhere on the list; it is, however, a real number between 0 and 1. This shows that our assumption that we have a correspondence between the set of all real numbers between 0 and 1 and the set of all natural numbers is in error: No such correspondence exists.

numbers. Cantor's proof shows that no matter what correspondence he chooses, at least one real number does not make it onto the list. The contradiction shows that our assumption that a one-to-one correspondence exists is in error. The set of real numbers is larger than the set of natural numbers. See the sidebar for details.

The proof that the set of real numbers is larger than the set of natural numbers has some surprising consequences. To appreciate one such consequence, recall that one can show that the set of all real algebraic numbers can be put into one-to-one correspondence with the set of all natural numbers. (Recall that a number is algebraic if it is the root of an algebraic equation.) The set of numbers

that are not algebraic is called transcendental, and so most numbers must be transcendental numbers; otherwise, the set of real numbers is no larger than the set of natural numbers. So although mathematicians can explicitly list relatively few transcendental numbers, it must be the case that “practically all” numbers are transcendental.

Although Cantor demonstrates that infinite sets of different sizes exist, the proof itself raises several interesting and important questions: Are there sets larger than the set of real numbers? How many “degrees of infinity” are there? Can these insights into infinite sets be used to create an arithmetic of infinite—technically called transfinite—numbers? The key to answering some of these questions involves the use of power sets.

A *power set* of a set S is the set of all subsets of S . Consider, for example, the set consisting of two letters, $\{a, b\}$. The power set of $\{a, b\}$ has four elements: $\{\{a\}, \{b\}, \{a, b\}, \emptyset\}$, where the symbol \emptyset represents the empty set, the set with no members. (The empty set is always a part of every set.) The power set of any nonempty set is larger than the set itself. This is more or less obvious for a finite set, since the power set must contain all single-element subsets of the original set as well as the empty set, but it is also true of infinite sets. If S represents an infinite set, then the power set of S always has a greater cardinality than the parent set from which it is derived. In other words we can never establish a one-to-one correspondence between any set and its power set.

Cantor shows that the power set of the natural numbers is the same size as the set of real numbers. Furthermore because the power set of the real numbers cannot be placed in one-to-one correspondence with the set of all real numbers—no set has the same cardinality as its power set—the power set of the real numbers is an even larger set than the set of real numbers. The power set of the power set of real numbers is still larger. The concept of the power set enables Cantor to show that there exist not one or two but infinitely many different sized infinite sets. He assigns a symbol to represent the size of each set. These symbols are examples of what are called cardinal numbers.

The cardinal numbers associated with infinite sets are a generalization of ordinary (finite) numbers, and understood

correctly they are a perfectly ordinary generalization. To understand the idea of transfinite numbers, thinking seriously about what an ordinary number is can be helpful. There is a difference, for example, between an instance of the number 3 and the number 3 itself. We are surrounded by instances of the number 3: A set of three lakes, a set of three cars, and a set of three women named Sue are all instances of the number 3. These sets are the same in the sense that we can choose any two of them and place the elements of these sets in one-to-one correspondence with each other. The correspondence is, in fact, the only characteristic that all instances of the number 3 have in common. Another way of describing this characteristic that all instances of a given number share is to apply the term *cardinality* to finite sets as well. Let S and T represent any two sets, finite or infinite. The cardinality of S , written $\text{Card}(S)$, tells us how many elements are in the set. The sets S and T are both instances of the same number if $\text{Card}(S) = \text{Card}(T)$; otherwise they are instances of different numbers.

The chief advantage of this approach is that it can be used word for word in describing “transfinite” cardinal numbers as well as finite ones. *Transfinite numbers* are numbers that represent the size of infinite sets. The natural numbers, the set of perfect squares, the set of rational numbers, and the set of algebraic numbers are all instances of the transfinite cardinal number that Cantor calls \aleph_0 (aleph null). So, for example, we can say that $\text{Card}(\{1, 2, 3, \dots\}) = \aleph_0$. The second transfinite number, represented, for example, by the set of all real numbers and the power set of the rational numbers, he calls \aleph_1 (aleph 1). By simply taking the power set of a set that is an instance of \aleph_1 , he is able to obtain an instance of \aleph_2 , and so on.

This is a new type of extension of the number system. Cantor has moved from finite numbers to transfinite numbers. He begins to inquire about the possibility of an arithmetic of transfinite numbers. He is successful in generalizing addition and multiplication into something that is now called cardinal arithmetic.

“The arithmetic of the infinite” may, at first, seem difficult to imagine. As in so much of mathematics, however, the key to new

ideas and new concepts lies in a reexamination of the familiar. To appreciate transfinite addition, we need to see how it is similar to ordinary addition. Let the plus symbol $+$ represent the operation of union of sets as well as its usual meaning of addition. Let A and B represent two sets. Then $A + B$ represents the set that is composed of all the elements of A and all the elements of B . For example if we let A represent the set $\{a, b\}$, and B represent the set $\{p, q, r\}$ then the set $A + B$ represents the set $\{a, b, p, q, r\}$.

Now consider the very simple equation $2 + 3 = 5$. We learned this equation early in grade school; now we only have to restate the equation in the language of sets. Let the set A be an instance of the number 2, for example, $A = \{a, b\}$, and let the set B be an instance of the number 3, for example, $B = \{p, q, r\}$. To understand Cantor's transfinite arithmetic, we just reinterpret $2 + 3 = 5$ as $\text{Card}(A) + \text{Card}(B) = \text{Card}(A + B)$. Now let x and y represent any numbers, finite or transfinite. Let X and Y represent instances of the cardinal numbers x and y , respectively. First we have $x + y = \text{Card}(X) + \text{Card}(Y)$. Because $\text{Card}(X) + \text{Card}(Y) = \text{Card}(X + Y)$, we conclude that $x + y = \text{Card}(X + Y)$: To find the sum of x and y we need only determine the cardinality of the union of X and Y . In this way Cantor extends the idea of addition from finite numbers to transfinite ones.

Multiplication can also be generalized to transfinite numbers. The inspiration for doing so also arises from reexamining the ordinary process of multiplication of ordinary numbers. To do this we use the multiplication symbol, \times , in a new way. Consider the following example: Let A and B again denote the sets $\{a, b\}$ and $\{p, q, r\}$, respectively. Define $A \times B$ to be the set of all ordered pairs that can be formed by using an element of A for the first coordinate and an element of B for the second coordinate. In this case $A \times B$ represents the following set: $\{(a, p), (a, q), (a, r), (b, p), (b, q), (b, r)\}$. Notice that when A has two elements and B has three elements $A \times B$ has six ($= 2 \times 3$) elements. It is just as easy to show that $B \times A$ also has six elements. This gives rise to the equation $\text{Card}(A) \times \text{Card}(B) = \text{Card}(A \times B)$, and this equation is taken to hold for every pair of sets, finite or infinite. Let x and y be two

numbers, either transfinite or finite. Let X and Y be instances of the numbers x and y . To multiply x by y we use the equation $x \times y = \text{Card}(X) \times \text{Card}(Y)$, and since $\text{Card}(X) \times \text{Card}(Y) = \text{Card}(X \times Y)$ we have our answer: $x \times y = \text{Card}(X \times Y)$.

Subtraction and division cannot be defined. The expression $\aleph_0 - \aleph_0$, for example, cannot be given any meaning. Neither, for example, can $\aleph_0 \div \aleph_0$. Nevertheless Cantor is able to imagine an arithmetic of the infinite that is, where it applies, an extension of ordinary finite arithmetic.

As do all good ideas Cantor's ideas raise many questions, some of which have not been answered. In particular Cantor's discovery of transfinite numbers leaves open the possibility that there are other, not-yet-discovered transfinite numbers. Cantor has demonstrated the existence of a sequence of transfinite numbers, that is, $\aleph_0, \aleph_1, \aleph_2, \dots$, but are there other transfinite numbers in addition to these? These numbers have to have the property that they fit in between Cantor's transfinite numbers. For example, does there exist a transfinite number that is strictly bigger than \aleph_0 and strictly smaller than \aleph_1 ? In this case an instance of this in between number is too big to be put into one-to-one correspondence with the set of natural numbers, and too small to be put into one-to-one correspondence with the set of real numbers. This problem, called the continuum hypothesis, is one of the famous problems in mathematics. It is easily generalized by asking whether there are

+	n	\aleph_0	\aleph_1
n	$2n$	\aleph_0	\aleph_1
\aleph_0	\aleph_0	\aleph_0	\aleph_1
\aleph_1	\aleph_1	\aleph_1	\aleph_1

\times	n	\aleph_0	\aleph_1
n	n^2	\aleph_0	\aleph_1
\aleph_0	\aleph_0	\aleph_0	\aleph_1
\aleph_1	\aleph_1	\aleph_1	\aleph_1

An addition table and a multiplication table that includes two transfinite numbers. The letter n represents an arbitrarily chosen natural number; and \aleph_0 and \aleph_1 represent the first two transfinite numbers.

other transfinite numbers that lie between each transfinite number in the sequence $\aleph_0, \aleph_1, \aleph_2, \dots$. This problem is called the generalized continuum hypothesis. These problems lie at the very foundation of mathematics.

Cantor's study of the infinite led to deep insights into the nature of infinite sets. It also led to new interpretations of the most elementary of arithmetic operations. Although his ideas remained controversial for some time, Cantor's set theory was soon embraced by some of the foremost mathematicians of his age because it held out hope of expressing all of mathematics in the language of sets. At the time it seemed at least possible that Cantor's ideas would lead to the axiomatization of all of mathematics. The result, however, was more complicated.

The Russell Paradox

Cantor's set theory offered a new way of thinking about mathematics. Points, lines, planes, numbers, arithmetic operations, and spaces—it seemed that everything could be expressed in the language of sets. One could talk about a collection of functions or one could describe a single function in the language of sets. Every function began to be understood as a special type of relation between a domain (a set) and a range (another set). This approach seems natural and uncontroversial to mathematics students today, but there was initially great controversy about Cantor's theory of sets.

The beauty of Cantor's creation is that it held out the promise of a new way of perceiving mathematics. With Cantor's abstract formulation of mathematics, mathematicians could concentrate on the relations between sets. This made certain mathematical facts that had been difficult to understand suddenly transparent. Cantor's work pushed back the frontiers of mathematics. Suddenly mathematicians could approach problems from a highly abstract and general point of view. Instead of working with individual functions, for example, they could work with a whole class of functions simultaneously. With this kind of generality many mathematicians hoped that eventually it would be possible



Bertrand Russell addressing an antinuclear weapons rally. In addition to his work as a mathematician and philosopher, Russell was a noted political activist. (Topham/The Image Works)

to discover and express the set-theoretic relations that lie at the foundation of mathematics. Cantor's discoveries, which were once perceived by many as abstract nonsense, describing nothing at all, now seemed to provide a language for describing everything. It was a tremendous success, and one that Cantor, toward the end of his life, must have recognized.

Cantor's last years were not easy, but despite periods of mental illness he continued to work on the theory of sets throughout his life. During this time he noticed a paradox associated with the theory of sets. This was a different kind of difficulty from the ones Cantor had faced before. Cantor's previous discoveries about sets had impressed many mathematicians as paradoxical, but the early definitions and theorems were not contradictory in

a logical sense. The early theorems were sometimes counterintuitive, but Cantor's arguments had prevailed because they were logically correct. Cantor's newly discovered paradox, however, revealed a strange and fundamental shortcoming in his formulation of set theory. This discovery, unlike his initial work on sets, did not attract much attention. For whatever reasons, ill health or uncertainty about the meaning of his discovery, Cantor allowed the paradox to remain largely unnoticed. Soon, however, other mathematicians had detected the same logical error in Cantor's formulation. One of the first and certainly one of the loudest mathematicians to detect the problem was the British philosopher, mathematician, and social activist Bertrand Russell (1872–1970).

As a boy Bertrand Russell was privately tutored. Growing up, he had lots of time to think and excellent teachers from whom to learn. As a consequence he began to grapple with difficult philosophical problems when he was still quite young. One of the problems that fascinated him as a teenager was the problem of certainty. Russell was less interested in what people “knew” to be true than he was in how people had begun to know it. (As an adult he wrote a number of books on the limits of understanding; *An Inquiry into Meaning and Truth* is one such work.) Russell eventually enrolled in Trinity College, Cambridge, where he distinguished himself through outstanding scholarship. Despite his enthusiasm for academic work, these interests did not define him.

Russell's interests always extended beyond scholarly activity. He always enjoyed writing nonscholarly as well as scholarly works and clearly enjoyed creating controversy. He worked hard at both. He wrote many fine books and articles, some for the specialist and some for the lay reader. Russell also held strong pacifist views and during World War I he was fined and jailed for six months for his activities against the war. Russell was undeterred. He used his time in jail to write more about mathematics. Russell's most famous mathematical work, the three-volume *Principia Mathematica*, written with his former tutor, the mathematician Alfred North Whitehead, was an attempt to derive

mathematics from first principles. This was an immense undertaking that neither author considered completely successful. It has been often said that though the book had two authors only one person has ever read the work cover to cover. In any case *Principia Mathematica* proved to be an influential text among logicians and mathematicians interested in the foundations of the subject. Russell continued to write about philosophy and mathematics throughout his life. He also remained active in social causes. In his last years he vociferously opposed the war in Vietnam.

With respect to his studies in set theory Russell took time to consider Cantor's ideas seriously. He was probably not the first to notice what has come to be known as the Russell paradox, but he almost certainly took more apparent delight in describing the paradox than anyone else. The Russell paradox, is a beautiful example of the type of conceptual difficulty that Russell enjoyed uncovering. The paradox revealed something quite unexpected about Cantor's version of set theory at a time when set theory was being incorporated into all of mathematics. The paradox also illustrates how careful we must be, even when doing something as simple as describing a set, if we hope to prevent logical contradictions.

To understand the Russell paradox, we begin with the problem of defining a set. In our early school years, when we first learn about sets, we often just list the elements in a set of interest. Rather than describing the set containing all U.S. states beginning with the letter *A*, we may, for example, simply list all the sets beginning with the letter *A*: Alaska, Alabama, Arizona, and Arkansas. Explicitly listing the elements in a set is fine for small sets, but for larger, more abstract sets, stating a condition that enables us to test whether any given object is a member of the set or not is wiser. The idea is simple enough: When presented with an object we simply compare the object against the condition of membership. If the object meets the condition for membership then it belongs to the set. If the object does not meet the condition for membership then it does not belong to the set. For example, suppose we let the letter *S* represent the

set of all pencils. “The set of all pencils” is the condition that defines membership in the set. Every pencil belongs to this set; at the same time our condition of membership excludes all frying pans, automobiles, and animals whose name begins with the letter a . This much is “obvious.” What is not so obvious is that this simple method of defining sets can also lead to logical contradictions.

Notice that one consequence of defining the set S as the set of pencils is that *only* pencils belong to S . In particular the set S , which is a set of pencils, is not itself a pencil, and so the set S is *not* an element of itself. If, however, we define the set T as the set of all objects that are *not* pencils then not only do frying pans, automobiles, and animals whose name begins with the letter a belong to T , but T also belongs to itself. The reason is simple: T meets the condition of membership. It is an object that is not a pencil. We can only conclude that the set T belongs to itself.

Defining a set through a simple membership condition as Cantor had done can lead to logical contradictions. To see the problem, let U represent the set consisting of all sets that are *not* members of themselves. In particular U contains S , the set of all pencils, which is defined in the previous paragraph. It may appear that U has a reasonable membership condition until we ask whether or not U is a member of itself. Since the answer to this question should be either yes or no, we consider each possible answer in turn:

1. If U is *not* a member of itself, then U meets the criterion for membership (in U) and so U must be a member of itself. This is a contradiction, of course. U cannot be both a member of itself and not a member of itself, and so our first conclusion is that we err when we assume that U is not a member of itself.
2. Suppose U is a member of itself, and recall that U is defined as the set containing all sets that are not members of themselves: If U belongs to U then U is *not* a member of itself—again a contradiction.

These two contradictions taken together show that U never existed. This is Russell's paradox, and here is an easy way to remember it: In a small town with one restaurant and one chef, the chef cooks for all those who do not cook for themselves. This statement makes perfect sense for everyone but the chef.

Russell's paradox can be used to disprove the existence of the concept of a universal set. The *universal set*, a set that mathematicians had originally found quite convenient to use, is the “set of all sets.” It is what we would get if we could form a gigantic set in which each element of the universal set is itself a set. The universal set, if it could exist, would contain *all* sets. If it existed, the universal set would contain the set of all pencils, the set of all frying pans, the set of all automobiles, and so on, but if the universal set existed, then it, too, would belong to itself, because it would meet the condition of membership: The universal set is a set. In fact the universal set would merely be one set among many that belong to the universal set, so the universal set would be a proper subset of itself. As a consequence we need to accept the existence of an even larger set that would contain all sets including the universal set, and therein lies the same contradiction. We *define* the universal set as the set that contains all sets. In particular there is no set larger than the universal set. Every set must belong to the universal set, but we have just shown that if the universal set existed then there must also exist an even larger set that contains it as a proper subset. The contradiction shows that we made an error in assuming the existence of the universal set. The only possible conclusion is that the universal set does not exist. No set can contain everything.

Resolving the Russell Paradox

The Russell paradox and other early set-theoretic paradoxes were associated with the existence of very large sets. One approach to placing set theory on a firm mathematical basis was to express the basic ideas of the theory as a collection of axioms chosen in such a way as to rule out the existence of very large sets.

Expressing a branch of mathematics through a collection of axioms was not a new idea. This was precisely the approach

adopted by the Greek geometer Euclid more than 2,000 years before Cantor's birth. In his most famous work *Elements*, an introduction to geometry, Euclid began by listing a set of axioms, postulates, and definitions. These statements were intended to define the subject matter of the book. The axioms and postulates described the basic properties of the geometry. (Today mathematicians make no distinction between axioms and postulates, and we also use the words interchangeably from now on.) Once the axioms were listed, the act of discovery in the geometry of Euclid consisted of making nonobvious logical deductions from the axioms. These logical deductions are called theorems. In a sense the theorems are nothing new. The information contained in the theorems was present all of the time in the axioms. A mathematician's job is simply to reveal the information that the axioms contain.

Reducing discovery to the act of making logical deductions from a set of axioms is what makes mathematics different from science. In science when there are several competing theories, the "correct" theory is determined by experiment. If experimental results conflict with mathematical reasoning, it is the mathematical reasoning that is modified or abandoned, not the experiment. Mathematicians, on the other hand, care nothing for experiment. It is irrelevant to their subject. The goal of the mathematician is to establish an unbroken chain of logical arguments from one proposition to the next. As long as the mathematics is logically coherent, nothing else matters.

The importance of the axiomatic method was recognized early. Its importance becomes apparent when one tries to do mathematics without axioms. Early in the history of mathematics, before Euclid, Greek mathematicians realized that working without axioms was not at all satisfactory. They recognized that it was not enough logically to deduce condition *C* from condition *B* and condition *B* from condition *A*. The problem is that this chain of logical inferences just keeps extending further and further back. In our example everything in the chain depends upon the truth of *A*. If condition *A* is true then so are conditions *B* and *C*, but if condition *A* is false then we have learned nothing at all about

either condition, *B* or *C*. The Greeks recognized that they were faced with an endless chain of implications, and their solution was to *begin* their deductions from a set of axioms whose truth was not open to question. These axioms cannot be proved within the confines of the subject because they define what the subject is. If we change the axioms we change the subject. The mathematical method is to begin with axioms and develop the subject from there.

Except for the geometry of Euclid not much thought was given to formal axiomatic systems until the 19th century, when several mathematicians, foremost among them the Russian Nikolai Ivanovich Lobachevsky (1793–1856), proposed a different set of axioms from the one adopted by Euclid. From these axioms Lobachevsky developed an exotic, new geometry. This geometry clashed with the common-sense notions that most people, even Lobachevsky, had about space, but Lobachevsky's geometry was sound in a mathematical sense, because it contained no internal contradictions. Mathematicians began to see the choice of one set of axioms over another as largely a matter of personal taste.

This is not to say that any set of statements qualifies as a set of axioms. Axioms must be free of contradictions in the sense that they cannot give rise to a statement that can be proved both true and false. They should be complete in the sense that those mathematical operations or results that mathematicians see as an integral part of their subject can be justified by the axioms. Finally, the



Nikolai Lobachevsky, sometimes called “the Copernicus of geometry,” changed the course of mathematics by creating and investigating the axioms for the first non-Euclidean geometry. (Library of Congress, Prints and Photographs Division [LC-USZ62-104326])

axioms should be free of repetition in the sense that one cannot prove one axiom as a consequence of the others. At first Lobachevsky's work was ignored, but his geometry and other subsequent attempts to axiomatize various branches of mathematics eventually attracted a lot of attention from other mathematicians. They realized that it was Lobachevsky's highly abstract, axiomatic approach to mathematics that held out the promise of a mathematics free of logical errors. It was with this understanding of mathematics that the German mathematician Ernest Zermelo (1871–1953) turned his attention to Cantor's set theory and the problems posed by Russell's paradox.

Zermelo was born in Berlin into comfortable surroundings. He was educated at universities in Berlin, Halle, and Freiburg, and he received a Ph.D. in mathematics from the University of Berlin. Initially Zermelo studied hydrodynamics, the mathematical modeling of fluid flow, but he soon turned his attention to set theory. As Cantor and Russell had, Zermelo had discovered what is now known as Russell's paradox, and he decided that set theory could be saved by axiomatizing the subject in such a way as to rule out the existence of very large sets. His solution was a set of seven axioms. From these seven axioms it was possible to deduce all of the major results of Cantor's set theory. The axioms were also chosen in such a way as to make deducing the paradoxes that had marred Cantor's original conception impossible. It was a great accomplishment.

One would think that Zermelo's accomplishment would have ended the controversy about the subject, but that was not the case. In order to accomplish his goal Zermelo needed an axiom that mathematicians everywhere now know as the axiom of choice. The *axiom of choice* is a statement about the existence of sets. It begins with a collection of sets, each of which is non-empty and disjoint from all the others. (A collection of sets is said to be disjoint when no two sets in the collection share a common element.) Essentially the axiom of choice states that given any collection of disjoint, nonempty sets, it is possible to create a new set that shares exactly one element in common with each of the sets in the original collection. It sounds innocent

enough, but the axiom of choice was the cause of many objections. The source of the objections was the idea that the axiom applied to *any* collection of nonempty sets. Even if we were considering as many sets as there are real numbers, for example, the axiom of choice assures us of the possibility of choosing elements from each set. (Of course with so many sets the word *choose* has to be used loosely.) This selection process is something that we might imagine doing, but many mathematicians objected to the axiom because it was not “constructive.” It was not always possible to devise an algorithm that would yield the necessary choices.

The axiom of choice allowed mathematicians to imagine the existence of an operation that had no basis in “reality,” but without the axiom of choice many of the most interesting results of set theory could not be derived. Mathematicians were faced with a new kind of choice: They could abandon new and useful mathematical ideas because those ideas were based on nonconstructive procedures, or they could adopt the axiom of choice and accept a new kind of mathematics, a mathematics that consisted of ideas and solutions that were occasionally nonconstructive. Even today a few mathematicians object enough to the axiom of choice to refuse to use it. Other mathematicians use the axiom of choice when they have to but avoid it when they can. Others embrace it wholeheartedly.

Zermelo made other important discoveries about set theory, but his career as a researcher was cut short by ill health. He began work at the University of Berlin and later worked at the University of Zurich. In 1916 he left Zurich as a result of poor health. He spent the next 10 years recovering from his illness. When he had recuperated, he found a position at the University of Freiburg in 1926, but he resigned in 1935 in a protest over Nazi policies. At the end of the war, in 1946 he applied to the University of Freiburg to be reinstated in his old position, and his request was granted. He died seven years later.

Cantor’s ideas enabled mathematicians to manipulate infinite sets successfully. This was extremely important to the development of all future mathematics, because so much of mathematics

is about the infinite. Furthermore his very general language of sets enabled mathematicians to perceive their work in new and more productive ways. Set theory became the language of mathematics. Many of the best mathematicians of the day tried to incorporate Cantor's ideas of the infinite into a new and more rigorous approach to math. Their efforts have been only partly successful. Nevertheless Cantor's approach to mathematics remains at the heart of the subject to this day.

12

CANTOR'S LEGACY

In the early years of the 20th century there were many mathematicians interested in the foundations of mathematics. In addition to Ernest Zermelo, two of the most prominent names associated with the attempt to understand mathematics at its most fundamental level were Bertrand Russell, the author of Russell's paradox, and the German mathematician David Hilbert (1862–1943). Hilbert received a Ph.D. from the University of Königsberg. (Königsberg is now called Kaliningrad and is part of Russia. At the time it was an important German cultural center.) He remained at Königsberg for several years after graduation as a teacher. Eventually he was hired at the University of Göttingen, where he remained for the rest of his working life.

Hilbert was a prolific mathematician, who made important contributions to many areas of mathematics and physics. He helped to establish the branch of mathematics now called functional analysis, in which functions are studied in the context of set theory. He also helped found the theory of infinite-dimensional spaces, which are now called Hilbert spaces. But Hilbert did more than find answers; he also famously posed questions. Today despite the many solutions that Hilbert found, he is best remembered for a series of questions that he challenged others to answer. In 1900 he posed 23 mathematical problems that he considered important to the future of the subject. Among these problems were questions that directly related to Cantor's work. Hilbert wanted to establish the truth of Cantor's continuum hypothesis, and he wanted to know whether it is possible to well-order any set. (We do not describe the concept of well-ordering here. We mention it because

this question was quickly answered by Ernest Zermelo, and Zermelo's proof required the axiom of choice.)

Hilbert's personal prestige and his acumen in choosing the problems seized the attention of mathematicians around the world, and a great deal of research during the 20th century was devoted to solving these problems. Many—but not all—of the problems were solved, and as mathematicians worked to unravel Hilbert's problems a great deal of interesting and useful mathematics was discovered along the way. Hilbert's choice of problems affected the direction of mathematical research—even long after his death.

Hilbert had always been interested in the fundamental ideas and relations that underpin the subject to which he devoted his life. It had long been recognized, for example, that there are logical flaws in Euclid's formulation of his geometry. One of Hilbert's early successes was to revisit the geometry of Euclid and provide a set of 21 axioms that provide a complete and logical foundation for this most ancient of all geometries. After 2,000 years of error Hilbert was the first to present a completely rigorous foundation for Euclid's geometry. For Hilbert this was just the beginning. He had a larger goal: He wanted to do for the entire field of mathematics what he had done for Euclidean geometry. His method was, in a sense, mechanical. Hilbert wanted to strip away all extraneous meanings from mathematics. In geometry for example the meanings that we attach to *point*, *line*, and *plane* can help us understand both the motivation for studying geometry and any applications that arise out of the study of the subject, but these meanings are irrelevant to geometry as a branch of mathematics.

In geometry mathematicians are interested only in the relationships *between* points, lines, and planes; they are not interested in what these words mean. Hilbert's goal was to treat all branches of mathematics in a similar way. He wanted to eliminate extraneous meanings and retain only abstract sets of axioms. From these axioms the theorems of the mathematical system were deduced. It was a purely formal procedure. Hilbert's concept of formal mathematics had much in common with board games. In a board game

we play by the rules. If we change the rules, we change the type of game that we are playing. In a board game our goal is to follow the rules, not create them. In mathematics once the axioms have been established all that remains is to deduce new theorems. The goal of every branch of mathematics is to use the axioms to prove new theorems. Mathematicians seldom create new axioms.

It may seem that once Hilbert had stripped all of the nonmathematical meaning from a mathematical system there would be nothing useful or interesting remaining. This is not true. Hilbert's formalized mathematical systems were like machines with the housing cut away to reveal the moving parts. We cut away the housing of a machine to see how it works, not to learn what the machine does. We can see what the machine does from the outside. To see *how* it does it, however, we need to look deeper. This was Hilbert's goal. No longer was there any question about whether the mathematical system under consideration was "true" or "correct" in the sense that it agreed with reality. Instead Hilbert's formalizations allowed us to see how—or even whether—a mathematical system generates good, logically consistent mathematics. Hilbert's goal was to look *inside* mathematics to discover the way it worked.

Keep in mind that in Hilbert's conception of formal mathematics the only question of interest is whether the theorems are direct, logical consequences of the axioms. This is a question best answered by examining the formal, abstract axioms that defined the subject. In this very rarefied atmosphere, the questions of the completeness and consistency of the axioms become central considerations. A set of axioms is *complete* if, given a statement in the mathematical system described by the axioms, the statement can always be proved either true or false. A complete set of axioms rules out the possibility of unprovable statements. A set of axioms is consistent provided every statement that is a consequence of the axioms is either true or false but not both true and false.

Hilbert began to search for criteria that would enable him to determine the completeness and consistency of particular sets of axioms. Hilbert, to be sure, never expected an easy answer to the questions of completeness and consistency. The fact that one has

*5241. $\vdash : \exists ! \alpha . \alpha \sim \epsilon 1 . \equiv . (\exists x, y) . x, y \in \alpha . x \neq y$

Dem.

$$\vdash . *24\cdot54 . \supset \vdash : \exists ! \alpha . \alpha \sim \epsilon 1 . \equiv : \alpha \neq \Lambda . \alpha \sim \epsilon 1 :$$

$$[*4\cdot56] \quad \equiv : \sim [\alpha \in 1 . \forall . \alpha = \Lambda] :$$

$$[*51\cdot236] \quad \equiv : \sim (\alpha \in 1 \cup t^{\epsilon} \Lambda) :$$

$$[*52\cdot4 . \text{Transp}] \quad \equiv : \sim [x, y \in \alpha . \supset_{x, y} x = y]$$

$$[*11\cdot52] \quad \equiv : (\exists x, y) . x, y \in \alpha . x \neq y : \supset \vdash . \text{Prop}$$

*5242. $\vdash : \alpha \in 1 . \supset : \exists ! \alpha \cap \beta . \equiv . \alpha \cap \beta \in 1$

Dem.

$$\vdash . *51\cdot31 . \quad \supset \vdash : \exists ! t^{\epsilon} x \cap \beta . \equiv . t^{\epsilon} x \cap \beta = t^{\epsilon} x :$$

$$[*20\cdot53] \quad \supset \vdash : \alpha = t^{\epsilon} x . \supset : \exists ! \alpha \cap \beta . \equiv . \alpha \cap \beta = t^{\epsilon} x :$$

$$[*10\cdot11\cdot28] \quad \supset \vdash : (\exists x) . \alpha = t^{\epsilon} x . \supset : (\exists x) : \exists ! \alpha \cap \beta . \equiv . \alpha \cap \beta = t^{\epsilon} x :$$

$$[*10\cdot37] \quad \supset : \exists ! \alpha \cap \beta . \supset : (\exists x) . \alpha \cap \beta = t^{\epsilon} x \quad (1)$$

$$\vdash . (1) . *52\cdot1 . \supset \vdash : \alpha \in 1 . \supset : \exists ! \alpha \cap \beta . \supset : \alpha \cap \beta \in 1 \quad (2)$$

$$\vdash . *52\cdot16 . \quad \supset \vdash : \alpha \cap \beta \in 1 . \supset : \exists ! \alpha \cap \beta \quad (3)$$

$$\vdash . (2), (3) . \quad \supset \vdash . \text{Prop}$$

*5243. $\vdash : \alpha \in 1 . \exists ! \alpha \cap \beta . \equiv . \alpha \in 1 . \alpha \cap \beta \in 1$ [*5242 . *532]

*5244. $\vdash : \alpha \in 1 . \supset : \exists ! \alpha \cap \beta . \equiv . \alpha C \beta . \equiv . \alpha \cap \beta = \alpha$

Dem.

$$\vdash . *51\cdot31 . \quad \supset \vdash : \exists ! t^{\epsilon} x \cap \beta . \equiv . t^{\epsilon} x C \beta :$$

$$[*13\cdot13 . \text{Exp}] \quad \supset \vdash : \alpha = t^{\epsilon} x . \supset : \exists ! \alpha \cap \beta . \equiv . \alpha C \beta :$$

$$[*10\cdot11\cdot23] \quad \supset \vdash : (\exists x) . \alpha = t^{\epsilon} x . \supset : \exists ! \alpha \cap \beta . \equiv . \alpha C \beta :$$

$$[*52\cdot1] \quad \supset \vdash : \alpha \in 1 . \supset : \exists ! \alpha \cap \beta . \equiv . \alpha C \beta \quad (1)$$

$$\vdash . (1) . *22\cdot621 . \supset \vdash . \text{Prop}$$

*5245. $\vdash :: \alpha, \beta \in 1 . \supset : \alpha C \beta \vee \gamma . \equiv : \alpha = \beta . v . \alpha C \gamma$

Dem.

$$\vdash . *51\cdot236 \frac{x, y, \gamma}{z, x, \beta} . \supset$$

$$\vdash :: x \in t^{\epsilon} y \vee \gamma . \equiv : x = y . v . x \in \gamma :$$

$$[*51\cdot2\cdot23] \quad \supset \vdash : t^{\epsilon} x C t^{\epsilon} y \vee \gamma . \equiv : t^{\epsilon} x = t^{\epsilon} y . v . t^{\epsilon} x C \gamma :$$

$$[*13\cdot21] \quad \supset \vdash : \alpha = t^{\epsilon} x . \beta = t^{\epsilon} y . \supset : \alpha C \beta \vee \gamma . \equiv : \alpha = \beta . v . \alpha C \gamma :$$

$$[*11\cdot11\cdot35] \quad \supset \vdash : (\exists x, y) . \alpha = t^{\epsilon} x . \beta = t^{\epsilon} y . \supset : \alpha C \beta \vee \gamma . \equiv : \alpha = \beta . v . \alpha C \gamma \quad (1)$$

$$\vdash . (1) . *52\cdot1 . \supset \vdash . \text{Prop}$$

*5246. $\vdash :: \alpha, \beta \in 1 . \supset : \alpha C \beta . \equiv . \alpha = \beta . \equiv . \exists ! (\alpha \cap \beta)$

Dem.

$$\vdash . *51\cdot2\cdot23 . \quad \supset : t^{\epsilon} x C t^{\epsilon} y . \equiv . t^{\epsilon} x = t^{\epsilon} y \quad (1)$$

$$\vdash . (1) . *13\cdot21 . \quad \supset : \alpha = t^{\epsilon} x . \beta = t^{\epsilon} y . \supset : \alpha C \beta . \equiv . \alpha = \beta \quad (2)$$

$$\vdash . (2) . *11\cdot11\cdot35 . *52\cdot1 . \supset : \alpha, \beta \in 1 . \supset : \alpha C \beta . \equiv . \alpha = \beta \quad (3)$$

$$\vdash . (3) . *52\cdot44 . \quad \supset \vdash . \text{Prop}$$

Text from Principia Mathematica showing the formal language developed by Whitehead and Russell to express their ideas (Courtesy of the Department of Special Collections, University of Vermont)

not yet found an unprovable statement does not mean that the axioms are complete. Nor does the fact that one has yet to find a statement that is both true and false mean that no inconsistency exists. Hilbert worked hard on the problems of completeness and consistency of sets of axioms. His plan of formalizing all of mathematics was not widely emulated, however.

Bertrand Russell took a different approach to understanding the foundations of mathematics. He wanted to show that pure mathematics is simply a branch of logic. Logic is that branch of knowledge that is concerned with the laws of rational thought. By beginning from first principles Russell hoped to derive mathematics from a sequence of carefully ordered logical steps and in this way prevent the logical paradoxes that had attracted his attention and the attention of others. This was one goal of the three-volume set *Principia Mathematica* that Russell and his old tutor Henry North Whitehead wrote together.

In *Principia* Russell and Whitehead hoped to recast mathematics. For them mathematics was not really about numbers; it was an exercise in logical thinking. Russell famously expressed his attitude about the whole field when he asserted that in mathematics one does not know what one is talking about nor whether what one is saying is true. Content and truth, Russell enjoyed asserting, are not the province of the mathematician. All that matters is whether the theorems are logically deduced from the axioms, and whether the axioms themselves are logically consistent with one another. Russell's paradox was a logical error. It has little to do with numbers. Because it represented a logical failure, one could correct it only by sharpening one's logical tools. This may sound relatively easy, but it is not. Russell and Whitehead spent years developing the necessary ideas and notation to express their ideas. They were not the first to attempt this, and they were heavily influenced by those who preceded them, but Russell and Whitehead were the most successful mathematicians of their time in expressing mathematical ideas in the language of logic.

As Hilbert was, Russell and Whitehead were in their own way attempting to create a formal system that would display the

workings of mathematics in a purely abstract setting. Over the course of their studies Russell and Whitehead returned to the set theory of Cantor and were able to deduce some of Cantor's main ideas by using their approach. They returned to the arithmetic of ordinary finite numbers and transfinite numbers, and they likewise reinterpreted these ideas in the language of logic. They did a lot, but there was a limit to their approach past which they could not go. Their reach exceeded their grasp. After three large volumes Whitehead and Russell stopped. By their own admission their attempt to exhibit even basic mathematics as a consistent formal system had not been entirely successful.

Kurt Gödel

Russell, Hilbert, Whitehead, Zermelo, and Cantor all sought to solidify the foundations of mathematics. But the work of all of these mathematicians was derailed in a surprising, even shocking way by the Austro-Hungarian-born mathematician Kurt Gödel (1906–78).

Gödel received a Ph.D. from the University of Vienna in 1929, and he remained at the university as a faculty member. During this period there was a great deal of political instability in Europe. The Nazi Party had seized power in Germany, and in Austria there was much sympathy for the ideas of national socialism as well. Gödel was not especially political. His life revolved around mathematics; he did his research and he published his results. As he gained prominence as a mathematician he was invited to travel to other countries to explain his discoveries. Whether Gödel initially noticed conditions in his country deteriorating—or, if he did notice, whether he much cared—is not clear from his actions. He was immersed in mathematics until the late 1930s, when a close friend and colleague of Gödel's was murdered by a student who was also a Nazi. This event profoundly affected Gödel. In 1938 he moved to the United States, where he joined the Princeton Institute for Advanced Study. He remained there for the rest of his working life.

It was in 1931, while Gödel was still in Vienna, that he published the discovery that turned mathematics on its head. His discovery was so unexpected that even now there are many mathematicians who are not sure what to make of it. Most mathematicians agree that Gödel's famous paper is important, and many assert that it should change ideas about mathematics, but few mathematicians can say how, or even whether, his result has affected their work. The uncertainty arises because of the peculiar nature of Gödel's discovery.

Gödel's discovery is summarized in his incompleteness theorems. As is much of the work that arose out of Cantor's early discoveries, Gödel's discovery is a negative result. He discovered that certain things are *not* possible. Essentially he discovered that Hilbert's intuitive belief—that it is possible to create a formal mathematics that is complete and consistent—was in error. The reason that Hilbert could not succeed in creating a complete, formal mathematics is not that he did not try hard enough or was not smart enough. It is that what Hilbert wanted to do is not



Kurt Gödel (second from right) receiving the Albert Einstein Award for Achievement in the Natural Sciences from Albert Einstein (Bettmann/CORBIS)

possible. It is not possible to create a formal system that is both complete and consistent.

In his paper Gödel discusses arithmetic, the simplest of all branches of mathematics. He considers the natural numbers together with the operations of addition and multiplication. He needs some logical context in which to place these “elementary” ideas, so he considers them in the context of Russell and Whitehead’s development of mathematics in their book *Principia Mathematica*. The result was Gödel’s incompleteness theorems. Essentially Gödel discovers that within any mathematical system there exist statements that cannot be either proved or disproved, and in particular he proved that this is true of arithmetic, the simplest and most important of all branches of mathematics.

Recall that a set of axioms is complete if every mathematical statement can be proved either true or false. It is consistent if no mathematical statement can be proved true *and* false. In his consideration of arithmetic Gödel manages to construct a statement about arithmetic that cannot be proved true and cannot be proved false. It is important to keep in mind that this does not imply that it is only Gödel who cannot prove the statement true or false. The content of the incompleteness theorems is that the statement cannot be proved true or false by anyone. It is impossible. This fact shows that the system is incomplete.

One may think that the reason Gödel’s statement cannot be proved either true or false lies with the axioms, that with a different or augmented set of axioms perhaps his statement can be proved true. It is true, of course, that we can always add axioms, but new axioms do not make the problem disappear. In fact there are always statements that are unprovable within a given set of axioms.

Gödel’s incompleteness theorem made quite an impression on his peers. Mathematicians had envisioned mathematics as a context in which truth—mathematical truth, at least—is really attainable. In fact many subscribed to the idea that only in mathematics can one derive (mathematical) truths that are certain and changeless. Gödel’s result *guarantees* the existence of statements

FORMAL LANGUAGES TODAY

The formalized mathematics of Hilbert and others never did catch on among mathematicians. This is due, in part, to the difficulty of learning the necessary notation and, in part, to the fact that they offer only a modest increase in clarity. It is also true that Gödel's incompleteness theorem demonstrates that the goal of the formal approach cannot, in any case, be completely realized. In retrospect it may appear that the highly formalized symbolic language pioneered in the work of Cantor, Hilbert, Russell, Whitehead, and others was a dead end, and mathematically this may prove to be the case. (We should, however, keep in mind that in mathematics the last chapter is never written.)

There are many others, however, who believe that formal languages are now some of the most important languages in our culture. Like the formal languages envisioned by Hilbert, these formal languages are not *about* anything. They have no meaning in themselves. They simply give us a method of describing relations that exist between unknown classes of objects. These formal languages are computer-programming languages, and they give us a way to describe relationships that exist between variables. The languages are a set of rules to manipulate variables. We may or may not know—or care—what the variables represent. The computer performs the manipulations without giving any consideration to whether the outcome is “true” or “correct,” and it displays results without regard to whether or not they agree with reality. The computer allows us to determine *only* what are direct, logical consequences of our assumptions, because the computer can only follow the rules that are encoded in the language. In this sense formal languages are now more important than ever. Some computer programs even resemble material in *Principia Mathematica* because of the abstract appearance of the symbols and the way that the language is expressed.

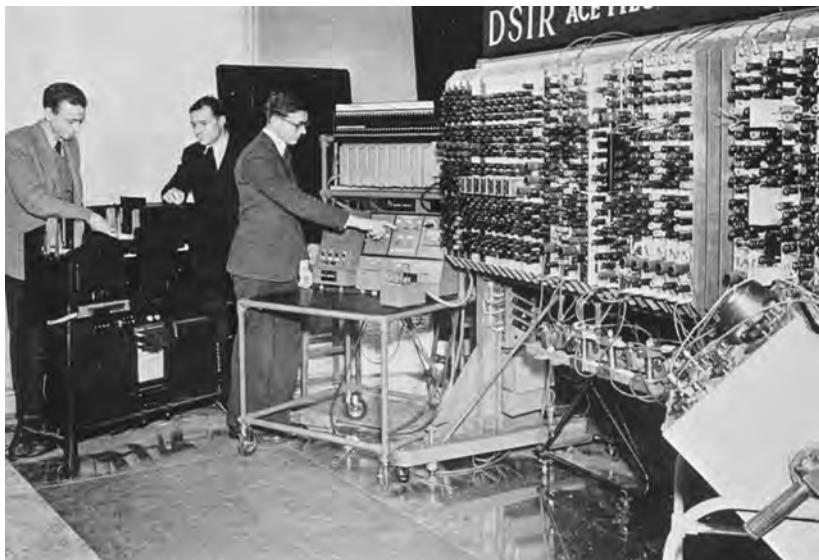
that are forever unprovable. Many prominent mathematicians found this news unsettling and even a little discouraging. Gödel shows that all of the work done by Zermelo, Hilbert, and others to establish mathematics on a solid foundation can be only partially successful. There are some goals that were permanently out of reach.

Alan Turing

Kurt Gödel's incompleteness theorems have a negative outcome. Gödel makes a statement about something that *cannot* be proved in arithmetic. His discovery has a counterpart in the field of computer science. This famous negative result was discovered by the British mathematician, cryptanalyst, and computer scientist Alan Turing (1912–54) during his early work on the foundations of computing.

Turing had a particularly creative and adaptive mind, and he had a special fondness for interdisciplinary problems. These are problems that span what many people consider separate branches of knowledge. Part of Turing's talent was his ability to draw on different ideas and combine them into something new. Today his most important work is usually thought to be his contribution to the then-emerging field of theoretical computer science. In fact his best-known work in computer science was done before there were any computers. He also contributed to the branch of knowledge called morphogenesis, the study of how form and structure develop in living organisms. During his own life he was best known for the role he played during World War II. Turing was recruited into a top-secret British project to break the secret German communication codes. These codes were, for the time, extremely sophisticated. Turing had a number of new ideas that enabled the Allies to decipher important German communications. This work was very important in helping the Allies win the war.

As might be expected from such a freewheeling intellect, Turing did not have an easy time at school. He paid little attention to the school curriculum, preferring to follow his own interests, and his teachers criticized him for his independent turn of mind. After secondary school he attended King's College, Cambridge University, where he studied mathematics. He then continued his education at the graduate school at King's College. During this time he began to write about mathematical logic, a topic that had grown in prominence since the publication of *Principia Mathematica*. Turing's research led him to conclude that to interpret mathematics simply as a branch of logic is unnecessarily



The first (1950) demonstration of the Pilot ACE computer. One of the world's earliest digital computers, the Pilot ACE was modeled on Alan Turing's design of the ACE, an even bigger, more powerful machine.
 (Science Museum, London/Topham-HIP/The Image Works)

narrow and restrictive. He believed it was more fruitful to consider mathematics from a variety of points of view. It was also during his time at King's College that he became familiar with Kurt Gödel's incompleteness results.

While he was in graduate school Turing began to think about a theoretical model for what we now call a computer, although it would be several more years before anything remotely similar to the machines with which we are familiar was developed. Turing's "programmable, multipurpose machine" was not a computer in the sense that Babbage understood the word. Babbage had tried to develop an actual programmable computing device. Turing, instead, described a sort of "thought experiment" that would help him express his ideas about the nature of computation. The Turing machine, as Turing's idea is now called, cannot be built—today or ever. Nevertheless Turing's ideas found wide application in the field of computer science then and now.

The Turing machine is a machine with unlimited memory. Turing also imagined a machine that could compute without interruption for days or even years at a time. In fact Turing imagined a machine that had *no limit whatsoever* on the amount of time that it could devote to a particular computation. This feature is important to keep in mind, because when a machine can compute forever, speed is no longer a consideration. In effect there is no rush, because the machine faces no time restrictions. A machine that can compute forever has only one limitation: its memory. But even that was not a problem because, as we have already mentioned, Turing also imagined that his machine had no limitations on its memory. Finally, Turing's machine was completely reliable. It made no mistakes and never broke down. It is easy to object to Turing's machine as impractical, but for what Turing had in mind his machine was better than a practical machine: It was a useful *idea*. Turing's idea helped make practical machines possible.

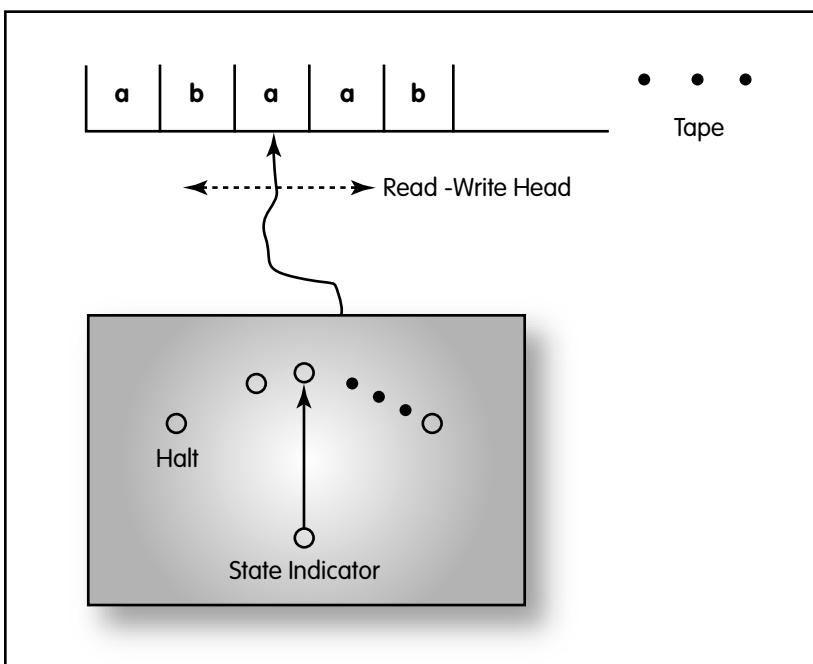


Diagram of a Turing machine. The dial indicates the state of the machine.

In Turing's concept the user communicated with the machine through a long tape. The tape was divided into squares. This is important because it shows that the machine responds to commands sequentially. Each square either is blank or contains a symbol printed on the square. Only a finite collection of symbols may be used to command the machine. At each step in the computing process the machine reads a square on the tape. As the symbol is read the machine responds accordingly. It may respond to the command by (1) erasing the square and leaving it blank, (2) erasing the square and printing a new symbol on the square, (3) leaving the square unaltered, or finally, (4) advancing the tape either forward or backward. The precise response of the machine is determined by the input symbol currently being read and the current state of the machine.

To understand what is meant by the *state of the machine*, think of the machine as consisting of a large collection of on–off switches. The state is determined by the configuration of the switches—that is, the pattern formed by the off and on switches. Eventually the altered tape—what we would call the output—is fed out of the machine. When the computer finishes its work—or, more precisely, *if* the computer finishes its work—it stops. The completed output tape represents the solution to the problem on which the computer was working. Turing's conceptual machine accepts input, manipulates information, and then produces output, and these functions are what make the Turing machine a sort of protocomputer.

Today most people have a very broad idea of what a computer can do. They think of it as a device for playing games, surfing the Internet, playing CDs and DVDs, and controlling a printer or other device. To Turing a computer is a device that computes. The Turing machine is no adding machine. It is a device that presents a mechanistic model of an activity that, up until that time, only humans have been capable of performing.

A device that can compute in the way a human can compute can do a lot, but what are its limits? Turing asked himself whether there is some task at which no computer can succeed. His question is so important to theoretical computer science that every

computer scientist recognizes Turing's problem by name. It is called the halting problem. A Turing machine that finishes its problem halts. A large problem or a slow Turing machine may require an enormous amount of time before it halts. The amount of time may be measured in days or years or even centuries. Turing, after all, is quite explicit: There are to be no limitations on the time it can devote to the solution of the problem. In theory there is no reason why a Turing machine might not compute for millions or even billions of years into the future. What Turing wants to know is whether there is a way to predict, using another Turing machine, whether a given Turing machine working on a given problem ever halts.

The halting problem cannot be answered experimentally. Running a machine and noting that it did not halt during the first 15 days or even 15 years of operation is no guarantee that it would not halt if allowed to operate longer. The halting problem can be answered only mathematically. Suppose that we have two Turing machines, machine A and machine B. Machine A has been programmed to compute the answer to some problem. We provide machine B with information about machine A as well as the problem that has been assigned to machine A to solve. We want machine B to tell us whether machine A will halt, that is, whether it will solve the problem, or cycle on endlessly without ever halting. What Turing discovers is that as a general rule, *machine B cannot determine whether or not machine A will halt*.

Turing's famous result encapsulates many ideas about the foundations of mathematics. To obtain his result, Turing uses some of the more advanced ideas pioneered by Cantor. His answer is expressed in the form of a contradiction, much in the way that Russell's paradox is expressed. Furthermore Turing's result contains an extension of Gödel's incompleteness results. An immediate consequence of Turing's initial discovery is that there is no set of axioms that enables a human to deduce whether an arbitrary program will halt or not. In other words the search for a conceptual framework to enable the user to distinguish the mathematically true from the mathematically false, leaving no ambiguity, has reached a dead end. Neither human nor machine can accomplish

this goal. The search for certainty has determined that there is no certainty to be had.

The concept of number is one of the most basic of all mathematical concepts. Numbers are, in many ways, the blocks with which mathematicians have built our mathematics. The process of developing our current conception of number has been difficult and long, because mathematicians had to shed their own preconceived notions of what a number is. They expanded their concept of number from positive rational numbers to negative rational numbers and the number 0. They moved from rational numbers to irrational numbers, and from irrational numbers to complex numbers. Finally, they moved from finite numbers to transfinite numbers. This required not just mathematical acumen but also the courage to state what in retrospect is clearly true and necessary at a time when it was perceived as obviously false and ridiculous. Throughout history people have mastered the basic arithmetic operations only with difficulty, but they have had no problem acquiring cultural prejudices about the meaning and uses of numbers. Despite the difficulties involved, the drive to expand the concept of number has been one of the most fruitful and prolonged inquiries in the history of the human race. These concepts have made science as well as mathematics possible.

It would be wrong to assert that all of the important discoveries about number and infinity have already been made. A brief review of the historical record shows that the pace of discovery has increased within the last few centuries and has been especially feverish in the last hundred or so years. There is no sign yet of a slowdown. Moreover some of the recent major discoveries, such as those of Gödel and Turing, are widely acknowledged as important but have yet to find their place in mainstream mathematical practice. Most professional mathematicians can describe the work of Turing and Gödel, but most are at a loss to explain how these fundamental ideas have affected their mathematical practice. Incompleteness and uncomputability are clearly important concepts, but to whom and why? Many mathematicians are still concerned with the implications of Gödel and Turing's discoveries.

Nor did the history of number end with Gödel and Turing. Other related questions that have attracted mathematicians' attention as potentially important remain unanswered. As of this writing, among these questions are Cantor's speculations about degrees of infinity. Other mathematicians are examining the number system for new types of structures; some mathematicians believe that there are aspects of our number system that are inherently random. The meaning and value of these new ideas are not yet clear, but it would not be surprising if, in future years, people's perceptions of the value of today's research changed radically. Nor is it only answers that remain to be discovered. In light of the history of mathematics it must certainly still be true that some of the major, foundational questions have yet to be discovered.

CHRONOLOGY

ca. 3000 B.C.E.

Hieroglyphic numerals are in use in Egypt.

ca. 2500 B.C.E.

Construction of the Great Pyramid of Khufu takes place.

ca. 2400 B.C.E.

An almost complete system of positional notation is in use in Mesopotamia.

ca. 1800 B.C.E.

The Code of Hammurabi is promulgated.

ca. 1650 B.C.E.

The Egyptian scribe Ahmes copies what is now known as the Ahmes (or Rhind) papyrus from an earlier version of the same document.

ca. 1200 B.C.E.

The Trojan War is fought.

ca. 740 B.C.E.

Homer composes the *Odyssey* and the *Iliad*, his epic poems about the Trojan War.

ca. 585 B.C.E.

Thales of Miletus carries out his research into geometry, marking the beginning of mathematics as a deductive science.

ca. 540 B.C.E.

Pythagoras of Samos establishes the Pythagorean school of philosophy.

ca. 500 B.C.E.

Rod numerals are in use in China.

ca. 420 B.C.E.

Zeno of Elea proposes his philosophical paradoxes.

ca. 399 B.C.E.

Socrates dies.

ca. 360 B.C.E.

Eudoxus, author of the method of exhaustion, carries out his research into mathematics.

ca. 350 B.C.E.

The Greek mathematician Menaechmus writes an important work on conic sections.

ca. 347 B.C.E.

Plato dies.

332 B.C.E.

Alexandria, Egypt, center of Greek mathematics, is established.

ca. 300 B.C.E.

Euclid of Alexandria writes *Elements*, one of the most influential mathematics books of all time.

ca. 260 B.C.E.

Aristarchus of Samos discovers a method for computing the ratio of the Earth-Moon distance to the Earth-Sun distance.

ca. 230 B.C.E.

Eratosthenes of Cyrene computes the circumference of Earth.

Apollonius of Perga writes *Conics*.

Archimedes of Syracuse writes *The Method*, *Equilibrium of Planes*, and other works.

206 B.C.E.

The Han dynasty is established; Chinese mathematics flourishes.

ca. C.E. 150

Ptolemy of Alexandria writes *Almagest*, the most influential astronomy text of antiquity.

ca. C.E. 250

Diophantus of Alexandria writes *Arithmetica*, an important step forward for algebra.

ca. 320

Pappus of Alexandria writes his *Collection*, one of the last influential Greek mathematical treatises.

415

The death of the Alexandrian philosopher and mathematician Hypatia marks the end of the Greek mathematical tradition.

ca. 476

The astronomer and mathematician Aryabhata is born; Indian mathematics flourishes.

ca. 630

The Hindu mathematician and astronomer Brahmagupta writes *Brahma-sphuta-siddhānta*, which contains a description of place-value notation.

641

The Library of Alexandria is burned.

ca. 775

Scholars in Baghdad begin to translate Hindu and Greek works into Arabic.

ca. 830

Mohammed ibn-Mūsā al-Khwārizmī writes *Hisāb al-jabr wa'l muqābala*, a new approach to algebra.

833

Al-Ma'mūn, founder of the House of Wisdom in Baghdad, Iraq, dies.

ca. 840

The Jainist mathematician Mahavira writes *Ganita Sara Samgraha*, an important mathematical textbook.

1071

William the Conqueror quells the last of the English rebellions.

1086

An intensive survey of the wealth of England is carried out and summarized in the tables and lists of the *Domesday Book*.

1123

Omar Khayyám, the author of *Al-jabr w'al muqābala* and the *Rubáiyát*, the last great classical Islamic mathematician, dies.

ca. 1144

Bhaskara II writes the *Lilavati* and the *Vija-Ganita*, two of the last great works in the classical Indian mathematical tradition.

ca. 1202

Leonardo of Pisa (Fibonacci), author of *Liber Abaci*, arrives in Europe.

1360

Nicholas Oresme, a French mathematician and Roman Catholic bishop, represents distance as the area beneath a velocity line.

1471

The German artist Albrecht Dürer is born.

1482

Leonardo da Vinci begins to keep his diaries.

ca. 1541

Niccolò Fontana, an Italian mathematician, also known as Tartaglia, discovers a general method for factoring third-degree algebraic equations.

1543

Copernicus publishes *De Revolutionibus*, marking the start of the Copernican revolution.

1545

Girolamo Cardano, an Italian mathematician and physician, publishes *Ars Magna*, marking the beginning of modern algebra. Later he publishes *Liber de Ludo Aleae*, the first book on probability.

ca. 1554

Sir Walter Raleigh, an explorer, adventurer, and amateur mathematician and patron of the mathematician Thomas Harriot, is born.

1579

François Viète, a French mathematician, publishes *Canon Mathematicus*, marking the beginning of modern algebraic notation.

1585

The Dutch mathematician and engineer Simon Stevin publishes “La disme.”

1609

Johannes Kepler, author of Kepler’s laws of planetary motion, publishes *Astronomia Nova*.

Galileo Galilei begins his astronomical observations.

1621

The English mathematician and astronomer Thomas Harriot dies. His only work, *Artis Analytiae Praxis*, is published in 1631.

ca. 1630

The French lawyer and mathematician Pierre de Fermat begins a lifetime of mathematical research. He is the first person to claim to have proved “Fermat’s last theorem.”

1636

Gérard (or Girard) Desargues, a French mathematician and engineer, publishes *Traité de la section perspective*, which marks the beginning of projective geometry.

1637

René Descartes, a French philosopher and mathematician, publishes *Discours de la méthode*, permanently changing both algebra and geometry.

1638

Galileo Galilei publishes *Dialogues Concerning Two New Sciences* while under arrest.

1640

Blaise Pascal, a French philosopher, scientist, and mathematician, publishes *Essai sur les coniques*, an extension of the work of Desargues.

1642

Blaise Pascal manufactures an early mechanical calculator, the Pascaline.

1648

The Thirty Years' War, a series of conflicts that involves much of Europe, ends.

1649

Oliver Cromwell takes control of the English government after a civil war.

1654

Pierre de Fermat and Blaise Pascal exchange a series of letters about probability, thereby inspiring many mathematicians to study the subject.

1655

John Wallis, an English mathematician and clergyman, publishes *Arithmetica Infinitorum*, an important work that presages calculus.

1657

Christiaan Huygens, a Dutch mathematician, astronomer, and physicist, publishes *De Ratiociniis in Aleae Ludo*, a highly influential text in probability theory.

1662

John Graunt, an English businessman and a pioneer in statistics, publishes his research on the London Bills of Mortality.

1673

Gottfried Leibniz, a German philosopher and mathematician, constructs a mechanical calculator that can perform addition, subtraction, multiplication, division, and extraction of roots.

1683

Seki Kōwa, Japanese mathematician, discovers the theory of determinants.

1684

Gottfried Leibniz publishes the first paper on calculus, *Nova Methodus pro Maximis et Minimis*.

1687

Isaac Newton, a British mathematician and physicist, publishes *Philosophiae Naturalis Principia Mathematica*, beginning a new era in science.

1693

Edmund Halley, a British mathematician and astronomer, undertakes a statistical study of the mortality rate in Breslau, Germany.

1698

Thomas Savery, an English engineer and inventor, patents the first steam engine.

1705

Jacob Bernoulli, a Swiss mathematician, dies. His major work on probability, *Ars Conjectandi*, is published in 1713.

1712

The first Newcomen steam engine is installed.

1718

Abraham de Moivre, a French mathematician, publishes *The Doctrine of Chances*, the most advanced text of the time on the theory of probability.

1743

The Anglo-Irish Anglican bishop and philosopher George Berkeley publishes *The Analyst*, an attack on the new mathematics pioneered by Isaac Newton and Gottfried Leibniz.

The French mathematician and philosopher Jean Le Rond d'Alembert begins work on the *Encyclopédie*, one of the great works of the Enlightenment.

1748

Leonhard Euler, a Swiss mathematician, publishes his *Introductio*.

1749

The French mathematician and scientist George-Louis Leclerc Buffon publishes the first volume of *Histoire naturelle*.

1750

Gabriel Cramer, a Swiss mathematician, publishes “Cramer’s rule,” a procedure for solving systems of linear equations.

1760

Daniel Bernoulli, a Swiss mathematician and scientist, publishes his probabilistic analysis of the risks and benefits of variolation against smallpox.

1761

Thomas Bayes, an English theologian and mathematician, dies. His “Essay Towards Solving a Problem in the Doctrine of Chances” is published two years later.

The English scientist Joseph Black proposes the idea of latent heat.

1762

Catherine II (Catherine the Great) is proclaimed empress of Russia.

1769

James Watt obtains his first steam engine patent.

1775

American colonists and British troops fight battles at Lexington and Concord, Massachusetts.

1778

Voltaire (François-Marie Arouet), a French writer and philosopher, dies.

1781

William Herschel, a German-born British musician and astronomer, discovers Uranus.

1789

Unrest in France culminates in the French Revolution.

1793

The Reign of Terror, a period of brutal, state-sanctioned repression, begins in France.

1794

The French mathematician Adrien-Marie Legendre (or Le Gendre) publishes his *Éléments de géométrie*, a text that influences mathematics education for decades.

Antoine-Laurent Lavoisier, a French scientist and discoverer of the law of conservation of matter, is executed by the French government.

1798

Benjamin Thompson (Count Rumford), a British physicist, proposes the equivalence of heat and work.

1799

Napoléon seizes control of the French government.

Caspar Wessel, a Norwegian mathematician and surveyor, publishes the first geometric representation of the complex numbers.

1801

Carl Friedrich Gauss, a German mathematician, publishes *Disquisitiones Arithmeticae*.

1805

Adrien-Marie Legendre, a French mathematician, publishes “Nouvelles méthodes pour la détermination des orbites des comètes,” which contains the first description of the method of least squares.

1806

Jean-Robert Argand, a French bookkeeper, accountant, and mathematician, develops the Argand diagram to represent complex numbers.

1812

Pierre-Simon Laplace, a French mathematician, publishes *Théorie analytique des probabilités*, the most influential 19th-century work on the theory of probability.

1815

Napoléon suffers final defeat at the battle of Waterloo.

Jean-Victor Poncelet, a French mathematician and the “father of projective geometry,” publishes *Traité des propriétés projectives des figures*.

1824

The French engineer Sadi Carnot publishes *Réflexions*, wherein he describes the Carnot engine.

Niels Henrik Abel, a Norwegian mathematician, publishes his proof of the impossibility of algebraically solving a general fifth-degree equation.

1826

Nikolay Ivanovich Lobachevsky, a Russian mathematician and “the Copernicus of geometry,” announces his theory of non-Euclidean geometry.

1828

Robert Brown, a Scottish botanist, publishes the first description of Brownian motion in “A Brief Account of Microscopical Observations.”

1830

Charles Babbage, a British mathematician and inventor, begins work on his analytical engine, the first attempt at a modern computer.

1832

János Bolyai, a Hungarian mathematician, publishes *Absolute Science of Space*.

The French mathematician Évariste Galois is killed in a duel.

1843

James Prescott Joule publishes his measurement of the mechanical equivalent of heat.

1846

The planet Neptune is discovered by the French mathematician Urbain-Jean-Joseph Le Verrier from a mathematical analysis of the orbit of Uranus.

1847

Georg Christian von Staudt publishes *Geometrie der Lage*, which shows that projective geometry can be expressed without any concept of length.

1848

Bernhard Bolzano, a Czech mathematician and theologian, dies. His study of infinite sets, *Paradoxien des Unendlichen*, is first published in 1851.

1850

Rudolph Clausius, a German mathematician and physicist, publishes his first paper on the theory of heat.

1851

William Thomson (Lord Kelvin), a British scientist, publishes “On the Dynamical Theory of Heat.”

1854

George Boole, a British mathematician, publishes *Laws of Thought*. The mathematics contained therein makes possible the later design of computer logic circuits.

The German mathematician Bernhard Riemann gives the historic lecture “On the Hypotheses That Form the Foundations of Geometry.” The ideas therein play an integral part in the theory of relativity.

1855

John Snow, a British physician, publishes “On the Mode of Communication of Cholera,” the first successful epidemiological study of a disease.

1859

James Clerk Maxwell, a British physicist, proposes a probabilistic model for the distribution of molecular velocities in a gas.

Charles Darwin, a British biologist, publishes *On the Origin of Species by Means of Natural Selection*.

1861

The American Civil War begins.

1866

The Austrian biologist and monk Gregor Mendel publishes his ideas on the theory of heredity in “Versuche über Pflanzenhybriden.”

1867

The Canadian Articles of Confederation unify the British colonies of North America.

1871

Otto von Bismarck is appointed first chancellor of the German Empire.

1872

The German mathematician Felix Klein announces his Erlanger Programm, an attempt to categorize all geometries with the use of group theory.

Lord Kelvin (William Thomson) develops an early analog computer to predict tides.

Richard Dedekind, a German mathematician, rigorously establishes the connection between real numbers and the real number line.

1874

Georg Cantor, a German mathematician, publishes “Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen,” a pioneering paper that shows that all infinite sets are not the same size.

1890

The Hollerith tabulator, an important innovation in calculating machines, is installed at the United States Census for use in the 1890 census.

1899

The German mathematician David Hilbert publishes the definitive axiomatic treatment of Euclidean geometry.

1900

David Hilbert announces his list of mathematics problems for the 20th century.

The Russian mathematician Andrey Andreyevich Markov begins his research into the theory of probability.

1901

Henri-Léon Lebesgue, a French mathematician, develops his theory of integration.

1905

Ernst Zermelo, a German mathematician, undertakes the task of axiomatizing set theory.

Albert Einstein, a German-born American physicist, begins to publish his discoveries in physics.

1906

Marian Smoluchowski, a Polish scientist, publishes his insights into Brownian motion.

1908

The Hardy-Weinberg law, containing ideas fundamental to population genetics, is published.

1910

Bertrand Russell, a British logician and philosopher, and Alfred North Whitehead, a British mathematician and philosopher, publish *Principia Mathematica*, an important work on the foundations of mathematics.

1914

World War I begins.

1917

Vladimir Ilyich Lenin leads a revolution that results in the founding of the Union of Soviet Socialist Republics.

1918

World War I ends.

The German mathematician Emmy Noether presents her ideas on the roles of symmetries in physics.

1929

Andrey Nikolayevich Kolmogorov, a Russian mathematician, publishes *General Theory of Measure and Probability Theory*, establishing the theory of probability on a firm axiomatic basis for the first time.

1930

Ronald Aylmer Fisher, a British geneticist and statistician, publishes *Genetical Theory of Natural Selection*, an important early attempt to express the theory of natural selection in mathematics.

1931

Kurt Gödel, an Austrian-born American mathematician, publishes his incompleteness proof.

The Differential Analyzer, an important development in analog computers, is developed at Massachusetts Institute of Technology.

1933

Karl Pearson, a British innovator in statistics, retires from University College, London.

1935

George Horace Gallup, a U.S. statistician, founds the American Institute of Public Opinion.

1937

The British mathematician Alan Turing publishes his insights on the limits of computability.

1939

World War II begins.

William Edwards Deming joins the United States Census Bureau.

1945

World War II ends.

1946

The Electronic Numerical Integrator and Calculator (ENIAC) computer begins operation at the University of Pennsylvania.

1948

While working at Bell Telephone Labs in the United States, Claude Shannon publishes “A Mathematical Theory of Communication,” marking the beginning of the Information Age.

1951

The Universal Automatic Computer (UNIVAC I) is installed at U.S. Bureau of the Census.

1954

FORmula TRANslator (FORTRAN), one of the first high-level computer languages, is introduced.

1956

The American Walter Shewhart, an innovator in the field of quality control, retires from Bell Telephone Laboratories.

1957

Olga Oleinik publishes “Discontinuous Solutions to Nonlinear Differential Equations,” a milestone in mathematical physics.

1964

IBM Corporation introduces the IBM System/360 computer for government agencies and large businesses.

1965

Andrey Nikolayevich Kolmogorov establishes the branch of mathematics now known as Kolmogorov complexity.

1966

A Programming Language (APL) computer language is implemented on the IBM System/360 computer.

1972

Amid much fanfare, the French mathematician and philosopher René Thom establishes a new field of mathematics called catastrophe theory.

1973

The C computer language, developed at Bell Laboratories, is essentially completed.

1975

The French geophysicist Jean Morlet helps develop a new kind of analysis based on what he calls “wavelets.”

1977

Digital Equipment Corporation introduces the VAX computer.

1981

IBM Corporation introduces the IBM personal computer (PC).

1989

The Belgian mathematician Ingrid Daubechies develops what has become the mathematical foundation for today’s wavelet research.

1991

The Union of Soviet Socialist Republics dissolves into 15 separate nations.

1995

The British mathematician Andrew Wiles publishes the first proof of Fermat’s last theorem.

Cray Research introduces the CRAY E-1200, a machine that sustains a rate of one teraflop (1 trillion calculations per second) on real-world applications.

JAVA computer language is introduced commercially by Sun Microsystems.

1997

René Thom declares the mathematical field of catastrophe theory “dead.”

2002

Experimental Mathematics celebrates its 10th anniversary. It is a refereed journal dedicated to the experimental aspects of mathematical research.

Manindra Agrawal, Neeraj Kayal, and Nitin Saxena create a brief, elegant algorithm to test whether a number is prime, thereby solving an important centuries-old problem.

2003

Grigory Perelman produces what may be the first complete proof of the Poincaré conjecture, a statement on the most fundamental properties of three-dimensional shapes.

GLOSSARY

algebraic equation an equation of the form $a_nx^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0 = 0$, where n can represent any natural number, x represents the variable raised to the power indicated, and a_j , which always denotes a rational number, is the coefficient by which x^j is multiplied

axiom a statement accepted as true that serves as a basis for deductive reasoning

axiomatize to develop a set of axioms for a mathematical system with the property that the system is complete and consistent

cardinal number the property that a set has in common with all sets equivalent to it. The cardinal number of a finite set is simply the number of elements in the set. For an infinite set the cardinal number is often used to identify a reference set with which the given set can be placed in one-to-one correspondence

coefficient a number or symbol used to multiply a variable

complete (set of axioms) a set of axioms defining a mathematical system with the property that every statement in the system can be proved either true or false

complex number any number of the form $a + bi$ where a and b are real numbers and i has the property that $i^2 = -1$

consistent (set of axioms) a set of axioms from which it is not possible to deduce a theorem that is both true and false

deduction a conclusion obtained by logically reasoning from general principles to particular statements

degree of an equation for an algebraic equation of one variable the largest exponent appearing in the equation

extraction of roots the procedure whereby given a number, here represented by the letter p , another number written $\sqrt[n]{p}$, the n th root of p , is computed. A number $\sqrt[n]{p}$ has the property that $(\sqrt[n]{p})^n = p$

fixed-point arithmetic any computational procedure in which one computes with numbers composed of a $t + s$ digits—where t digits are used to represent the integer part of each number and s digits are used to represent the fractional part

floating-point arithmetic any arithmetic procedure performed with numbers that are in normalized floating-decimal form. In base 10 that means any number, p , represented in the form $p = m \times 10^n$ where n is an integer and m is a number consisting of a fixed number of digits and $1 \leq m < 10$. Similar representations exist for other bases

halting problem a problem, conceived by Alan Turing, that cannot be solved by any type of automatic computation. The problem itself involves predicting whether or not a computer will reach the end of a particular type of computation (and halt) or simply continue to compute forever

incomplete (set of axioms) a set of axioms describing a mathematical system in which there exists a statement within the system that cannot be proved either true or false as a logical consequence of the axioms

infinite set a set that can be placed in one-to-one correspondence with a proper subset of itself

integer any number belonging to the set $\{0, -1, 1, -2, 2, -3, 3, \dots\}$

irrational number any real number that cannot be written as a/b where a and b are integers and b is not 0

method of exhaustion the proposition in Greek geometry that given any magnitude M one can, by continually reducing its size by at least half, make the magnitude as small as desired. Given a “small” positive number, usually denoted by the Greek letter ϵ (epsilon), and a number r such that $0 < r < 1/2$, then $M \times r^n < \epsilon$ provided that n is a sufficiently large natural number. This proposition formed the basis for the Greek analog to calculus

one-to-one correspondence the pairing of elements between two sets, A and B , so that each element of A is paired with a unique element of B and to each element of B is paired a unique element of A

perfect square an integer p with the property that there exists another integer q such that $q^2 = p$

positional numeration a system of notation in which each number is represented by a string of digits and the value of a digit depends on its position in the string

power set given a set S , the power set of S is defined as the set whose elements are all of the subsets of S

proper subset given two sets A and B , B is a proper subset of A if every element of B belongs to A but not every element of A belongs to B

quadratic equation an algebraic equation in which the highest exponent appearing is of degree 2

quadratic formula a mathematical formula for computing the roots of any second-degree algebraic equation by using the coefficients that appear in the equation

rational number any number of the form a/b where a and b are integers and b is not 0

real number any number that corresponds to a point on the real number line

root (of an equation) any number that satisfies an algebraic equation

Russell paradox a logical paradox that arises when a set is defined so that it is an element of itself. The Russell paradox shows that the “set of all sets” does not exist

set a collection of numbers, points, or other elements

sexagesimal base 60

subset given two sets, A and B , B is said to be a subset of A if every element of B also belongs to A

theorem a statement that is deduced from the axioms that define a given mathematical system

transcendental number a number that is not the root of any algebraic equation

transfinite number the cardinal number of an infinite set

Turing machine a theoretical machine used to pose and solve certain problems in computer science

unit fraction a fraction of the form $1/a$ where a is any integer except 0

vigesimal base 20

FURTHER READING

MODERN WORKS

Aczel, Amir D. *The Mystery of the Aleph: Mathematics, the Kaballah, and the Human Mind*. New York: Four Walls Eight Windows, 2000. This is the story of how mathematicians developed the concept of infinity. It also covers the religious motivations of Georg Cantor, the founder of modern set theory.

Adler, Irving. *Thinking Machines, a Layman's Introduction to Logic, Boolean Algebra, and Computers*. New York: John Day, 1961. This old book is still the best nontechnical introduction to computer arithmetic. It begins with fingers and ends with Boolean logic circuits.

Boyer, Carl B., and Uta C. Merzbach. *A History of Mathematics*. New York: John Wiley & Sons, 1991. Boyer was one of the preeminent mathematics historians of the 20th century. This work contains much interesting biographical information. The mathematical information assumes a fairly strong background of the reader.

Bruno, Leonard C. *Math and Mathematicians: The History of Mathematics Discoveries around the World*, 2 vols. Michigan: U.X.L, 1999. Despite its name there is little mathematics in this two-volume set. What you find is a very large number of brief biographies of many individuals who were important in the history of mathematics.

Bunt, Lucas Nicolaas Hendrik, Phillip S. Jones, Jack D. Bedient. *The Historical Roots of Elementary Mathematics*. Englewood Cliffs, N.J.: Prentice Hall, 1976. A highly detailed examination—complete with numerous exercises—of how ancient cultures added, subtracted, divided, multiplied, and reasoned.

Courant, Richard, and Herbert Robbins. *What Is Mathematics? An Elementary Approach to Ideas and Mathematics*. New York: Oxford University Press, 1941. A classic and exhaustive answer to the

question posed in the title. Courant was an influential 20th-century mathematician.

Danzig, Tobias. *Number, the Language of Science*. New York: Macmillan, 1954. First published in 1930, this book is painfully elitist; the author's many prejudices are on display in every chapter. Yet it is one of the best nontechnical histories of the concept of number ever written. Apparently it was also Albert Einstein's favorite book on the history of mathematics.

Davis, Phillip J. *The Lore of Large Numbers*. New York: Random House, 1961. An excellent overview of numbers, ways they are written, and ways they are used in science.

Dewdney, Alexander K. *200% of Nothing: An Eye-Opening Tour through the Twists and Turns of Math Abuse and Innumeracy*. New York: John Wiley & Sons, 1993. A critical look at ways mathematical reasoning has been abused to distort truth.

Eastaway, Robert, and Jeremy Wyndham. *Why Do Buses Come in Threes? The Hidden Mathematics of Everyday Life*. New York: John Wiley & Sons, 1998. Nineteen lighthearted essays on the mathematics underlying everything from luck to scheduling problems.

Eves, Howard. *An Introduction to the History of Mathematics*. New York: Holt, Rinehart & Winston, 1953. This well-written history of mathematics places special emphasis on early mathematics. It is unusual because the history is accompanied by numerous mathematical problems. (The solutions are in the back of the book.)

Gardner, Martin. *The Colossal Book of Mathematics*. New York: Norton, 2001. Martin Gardner had a gift for seeing things mathematically. This "colossal" book contains sections on geometry, algebra, probability, logic, and more.

_____. *Logic Machines and Diagrams*. Chicago: University of Chicago Press, 1982. An excellent book on logic and its uses in computers.

Guillen, Michael. *Bridges to Infinity: The Human Side of Mathematics*. Los Angeles: Jeremy P. Tarcher, 1983. This book consists of an engaging nontechnical set of essays on mathematical topics,

including non-Euclidean geometry, transfinite numbers, and catastrophe theory.

Hoffman, Paul. *Archimedes' Revenge: The Joys and Perils of Mathematics*. New York: Ballantine, 1989. A relaxed, sometimes silly look at an interesting and diverse set of math problems ranging from prime numbers and cryptography to Turing machines and the mathematics of democratic processes.

Joseph, George G. *The Crest of the Peacock: The Non-European Roots of Mathematics*. Princeton, N.J.: Princeton University Press, 1991. One of the best of a new crop of books devoted to this important topic.

Kaplan, Robert. *The Nothing That Is: A Natural History of Zero*. New York: Oxford University Press, 1999. Zero was one of the great conceptual breakthroughs in the history of mathematics. This is a lively history devoted to that one concept.

Kasner, Edward, and James R. Newman. "Paradox Lost and Paradox Regained." In *The World of Mathematics*. Vol. 3. New York: Dover Publications, 1956. A discussion of many common mathematical paradoxes and their logical resolutions.

Kline, Morris. *Mathematics for the Nonmathematician*. New York: Dover Publications, 1985. An articulate, not very technical overview of many important mathematical ideas.

_____. *Mathematics in Western Culture*. New York: Oxford University Press, 1953. An excellent overview of the development of Western mathematics in its cultural context, this book is aimed at an audience with a firm grasp of high school-level mathematics.

Maor, Eli. *e: The Story of a Number*. New York: Oxford University Press, 1999. The transcendental number called *e* plays a peculiar and very important role in calculus and the mathematics that grew out of calculus. This is an entire book dedicated to telling the story of *e*.

McLeish, John. *Number*. New York: Fawcett Columbine, 1992. A history of the concept of number from Mesopotamian to modern times.

Niven, Ivan Morton. *Numbers, Rational and Irrational*. New York: Random House, 1961. Aimed at older high school students, this is a challenging mathematical presentation of our number system.

Pappas, Theoni. *The Joy of Mathematics*. San Carlos, Calif.: World Wide/Tetra, 1986. Aimed at a younger audience, this work searches for interesting applications of mathematics in the world around us.

Pierce, John R. *An Introduction to Information Theory: Symbols, Signals and Noise*. New York: Dover Publications, 1961. Despite the sound of the title, this is not a textbook. Pierce, formerly of Bell Laboratories, describes how (among other topics) numbers and text are digitally encoded for transmission and storage—a lucid introduction to a very important topic.

Reid, Constance. *From Zero to Infinity: What Makes Numbers Interesting*. New York: Thomas Y. Crowell, 1960. A well-written overview of numbers and the algebra that stimulated their development.

Seife, Charles. *Zero: The Biography of a Dangerous Idea*. New York: Viking Press, 2000. A book dedicated to one of the most productive ideas in the history of mathematics.

Smith, David E., and Yoshio Mikami. *A History of Japanese Mathematics*. Chicago: Open Court, 1914. Copies of this book are still around, and it is frequently quoted. The first half is an informative nontechnical survey. The second half is written more for the expert.

Stewart, Ian. *From Here to Infinity*. New York: Oxford University Press, 1996. A well-written, very readable overview of several important contemporary ideas in geometry, algebra, computability, chaos, and mathematics in nature.

Swetz, Frank J., editor. *From Five Fingers to Infinity: A Journey through the History of Mathematics*. Chicago: Open Court, 1994. This is a fascinating, though not especially focused, look at the history of mathematics. Highly recommended.

Tabak, John. *Algebra. History of Mathematics*. New York: Facts On File, 2004. More information about how the concept of number and ideas about the nature of algebra evolved together.

Thomas, David A. *Math Projects for Young Scientists*. New York: Franklin Watts, 1988. This project-oriented text gives an introduction to several historically important geometry problems.

Zippin, Leo. *The Uses of Infinity*. New York: Random House, 1962. Contains lots of equations—perhaps too many for the uninitiated—but none of the equations is very difficult. The book is worth the effort needed to read it.

ORIGINAL SOURCES

Reading the discoverer's own description can sometimes deepen our appreciation of an important mathematical discovery. Often this is not possible, because the description is too technical. Fortunately there are exceptions. Sometimes the discovery is accessible because the idea does not require a lot of technical background to be appreciated. Sometimes the discoverer writes a nontechnical account of the technical idea that she or he has discovered. Here are some classic papers:

Ahmes. *The Rhind Mathematical Papyrus: Free Translation, Commentary, and Selected Photographs, Transcription, Literal Translations*. Translated by Arnold B. Chace. Reston, Va.: National Council of Teachers of Mathematics, 1979. This is a translation of the longest and best of extant Egyptian mathematical texts, the Rhind papyrus (also known as the Ahmes papyrus). It provides insight into the types of problems and methods of solution known to one of humanity's oldest cultures.

Galilei, Galileo. *Dialogues Concerning Two New Sciences*. Translated by Henry Crew and Alphonso de Salvio. New York: Dover Publications, 1954. Galileo's discussion of infinite sets begins on page 31 of this edition. It is fascinating to read his grappling with this difficult subject, and as he struggles with the idea of infinity he sees further into the subject than any of his predecessors.

Hahn, Hans. Infinity. In *The World of Mathematics*. Vol. 3, edited by James Newman. New York: Dover Publications, 1956. A thoughtful exposition of some of the fundamental ideas involved in the study of infinite sets by one of the great mathematicians of the first half of the 20th century.

Hardy, Godfrey H. *A Mathematician's Apology*. Cambridge, England: Cambridge University Press, 1940. Hardy was an excellent mathematician and a good writer. In this oft-quoted and very brief book Hardy seeks to explain and sometimes justify his life as a mathematician.

Russell, Bertrand. Definition of Number. In *The World of Mathematics*. Vol. 1, edited by James R. Newman. New York: Dover Publications, 1956. Russell was one of the great thinkers of the 20th century and a pioneer in mathematics. Here he describes for a general audience contemporary insights into the nature of number.

———. *Mathematics and the Metaphysicians*. In *The World of Mathematics*. Vol. 3, edited by James Newman. New York: Dover Publications, 1956. An introduction to the philosophical ideas on which mathematics is founded written by a major contributor to this field.

Stevin, Simon. *Principal Works*, edited by Ernest Crane, E. J. Dijksterhuis, A. D. Fokker, R. J. Forbes, A. Pannekock, A. Romein-Verschoor, W. H. Schukking, D. J. Struik. Translated by C. Dikshoorn. 5 vols. Amsterdam: C. V. Swets and Zeitlinger, 1955–1966. Every history of mathematics devotes space to Simon Stevin, but unlike those of Galileo and Kepler, Stevin's writings are difficult to find. This very readable translation is available in some larger libraries.

Turing, Alan M. “Can a Machine Think?” In *The World of Mathematics*. Vol. 4, edited by James R. Newman. New York: Dover, 1956. In many ways modern theoretical computer science begins with the work of Turing. This early nontechnical article deals with the idea of artificial intelligence. Some of Turing's vocabulary is a little old-fashioned, but the ideas are important and relevant.

INTERNET RESOURCES

Athena Earth and Space Science for K–12. Available on-line. URL: <http://inspire.ospi.wednet.edu:8001/>. Updated May 13, 1999. Funded by the National Aeronautics and Space Administration's (NASA's) Public Use of Remote Sensing Data, this site contains many interesting applications of mathematics to the study of natural phenomena.

Boyne, Anne. Papers on History of Science. http://nti.educa.rcanaria.es/penelope/uk_confboyne.htm#particular. Downloaded June 2, 2003. This is a very detailed and interesting paper devoted to the history of negative numbers. It is well worth reading.

The Eisenhower National Clearinghouse for Mathematics and Science Education. Available on-line. URL: <http://www.enc.org/>. Downloaded June 2, 2003. As its name implies, this site is a clearinghouse for a comprehensive set of links to interesting sites in math and science.

Electronic Bookshelf. Available on-line. URL: <http://hilbert.dartmouth.edu/~matc/eBookshelf/art/index.html>. Updated May 21, 2002. This site is maintained by Dartmouth College. It is both visually beautiful and informative, and it has links to many creative presentations on computer science, the history of mathematics, and mathematics. It also treats a number of other topics from a mathematical perspective.

Eric Weisstein's World of Mathematics. Available on-line. URL: <http://mathworld.wolfram.com/>. Updated April 10, 2002. This site has brief overviews of a great many topics in mathematics. The level of presentation varies substantially from topic to topic.

Faber, Vance, Bonnie Yantis, Mike Hawrylycz, Nancy Casey, Mike Fellows, Mike Barnett, Gretchen Wissner. This is MEGA Mathematics! Available on-line. URL: <http://www.c3.lanl.gov/mega-math>. Downloaded June 2, 2003. Maintained by the Los Alamos National Laboratories, one of the premier scientific establishments in the world, this site has a number of unusual offerings. It is well worth a visit.

Fife, Earl, and Larry Husch. Math Archives. "History of Mathematics." Available on-line. URL: <http://archives.math.utk.edu/topics/history.html>. Updated January 2002. Information on mathematics, mathematicians, and mathematical organizations.

Gangolli, Ramesh. *Asian Contributions to Mathematics*. Available on-line. URL: <http://www.pps.k12.or.us/depts-c/mc-me/be-asma.pdf>. Downloaded June 2, 2003. As its name implies, this well-written on-line book focuses on the history of mathematics in Asia and its effect on the world history of mathematics. It also includes information on the work of Asian Americans, a welcome contribution to the field.

Goldsmith, Asaf, and Atsushi Akera. "John W. Mauchly and the Development of the ENIAC Computer," Van Pelt Library, University of Pennsylvania. Available on-line. URL: <http://oldsite.library.upenn.edu/exhibits/rbm/mauchly/jwmintro.html>. Updated April 23, 2003. This site, maintained by the University of Pennsylvania, has a number of interesting pictures of the ENIAC computer and a compelling story to tell.

Heinlow, Lance, and Karen Pagel. "Math History." Online Resource. Available on-line. URL: <http://www.amatyc.org/OnlineResource/index.html>. Updated May 14, 2003. Created under the auspices of the American Mathematical Association of Two-Year Colleges, this site is a very extensive collection of links to mathematical and math-related topics.

The Math Forum @ Drexel. The Math Forum Student Center. Available on-line. URL: <http://mathforum.org/students/>. Updated June 2, 2003. Probably the best website for information about the mathematics that students encounter in their school-related studies. You will find interesting and challenging problems and solutions for students in grades K–12 as well as a fair amount of college-level information.

Melville, Duncan J. Mesopotamian Mathematics. Available on-line. URL: <http://it.stlawu.edu/ca.dmelvill/mesomath/>. Updated March 17, 2003. This creative site is devoted to many aspects of Mesopotamian mathematics. It also has a link to a "cuneiform calculator," which can be fun to use.

O'Connor, John L., and Edmund F. Robertson. The MacTutor History of Mathematics Archive. Available on-line. URL: <http://www.gap.dcs.st-and.ac.uk/~history/index.html>. Updated May 2003. This is a valuable resource for anyone interested in learning more about the history of mathematics. It contains an extraordinary collection of biographies of mathematicians in different cultures and times. In addition it provides information about the historical development of certain key mathematical ideas.

Siegrist, Kyle. Dice Experiment. Available on-line. URL: <http://www.math.uah.edu/psol/applets/DiceExperiment.html>. Downloaded June 20, 2003. This is a very thoughtfully designed demonstration of some very important ideas in the theory of probability. Highly recommended.

PERIODICALS, THROUGH THE MAIL AND ON-LINE

+Plus

URL: <http://pass.maths.org.uk>

A site with numerous interesting articles about all aspects of high school math. They send an email every few weeks to their subscribers to keep them informed about new articles at the site.

Function

Business Manager

Department of Mathematics and Statistics

Monash University

Victoria 3800

Australia

function@maths.monash.edu.au

Published five times per year, this refereed journal is aimed at older high school students.

The Math Goodies Newsletter

<http://www.mathgoodies.com/newsletter/>

A popular, free e-newsletter that is sent out twice per month.

Parabola: A Mathematics Magazine for Secondary Students

Australian Mathematics Trust

University of Canberra

ACT 2601

Australia

Published twice a year by the Australian Mathematics Trust in association with the University of New South Wales, *Parabola* is a source of short high-quality articles on many aspects of mathematics. Some back issues are also available free on-line. See URL: <http://www.maths.unsw.edu.au/Parabola/index.html>.

Pi in the Sky

<http://www.pims.math.ca/pi/>

Part of the Pacific Institute for the Mathematical Sciences, this high school mathematics magazine is available over the Internet.

Scientific American

415 Madison Avenue

New York, NY 10017

A serious and widely read monthly magazine, *Scientific American* regularly carries high-quality articles on mathematics and mathematically intensive branches of science. This is the one “popular” source of high-quality mathematical information that you will find at a newsstand.

INDEX

Italic page numbers indicate illustrations.

A

- abacists 39–40
- abacus 39
- aboriginal Australians, absence of counting system 9
- absolute value 105
- addition
 - positive rational numbers as closed system under 69, 71
 - real numbers as closed system under 92
 - transfinite 152, 153
- Ahmes papyrus 21
 - hieratic script in 22
 - problem from 24–25
- Alembert, Jean Le Rond d' 100–101
 - accomplishments of 101
 - and algebraic numbers, contribution to understanding of 102
- aleph null 151
- Alexander the Great (king of Macedonia) 16
- algebra, fundamental theorem of 94
- algebraic equations 85–86
 - and algebraic numbers, relationship between 94
 - degree of 86
 - and number of roots 94

- fourth-degree, Ferrari's algorithm for 90, 91
- Girard's solutions to 93
- Renaissance algorithms for 86, 87
- roots of 87, 107
 - Descartes on 95
 - Girard on 94
 - Wallis on 95–96
- second-degree 86, 87
 - geometric representation of roots of 95–96
- third- and fourth-degree, new numbers discovered through 87
- third-degree, Tartaglia's algorithm for 89–90
- algebraic numbers
- d'Alembert's contribution to understanding of 102
- and algebraic equations, relationship between 94
 - definition of 88, 107
 - expansion of concept of number to include 92
 - and natural numbers, correspondence between 146–147
- algorists 39
- analytical engine 50–54, 51

- approximation
 - computers and 56–57, 72
 - Mesopotamian algorithm for 71, 72
 - Stevin on 41
- Arabic numerals, introduction in Europe 35
- Argand, Jean-Robert 103
- Argand diagram 103, 104, 105–106
- Aristotle, on infinity 128
- arithmetic
 - basic operations of 69
 - Gödel's consideration of 172
 - of transfinite numbers 151–153, 153
- Arithmetica Infinitorum* (Wallis) 95
- Ars Magna* (Cardano) 88, 91–92
- The Art of Nombrystng* (Sacrobosco) 38–39
- Aryabhata I (Hindu mathematician) 82
- astronomy
 - Galileo and 130
 - Indian 78, 82
 - Mayan 27, 29
- axiom(s) 159
 - completeness and consistency of, questions of 167–169
 - criteria for 161–162

Euclid's 160
 Hilbert's 166
 importance of
 160–161
 Lobachevsky's 161
 logical deductions
 from 160
 Zermelo's 162–163
 axiom of choice
 162–163

B

Babbage, Charles, analytical engine of
 50–54, 51
 Babylon, Hanging Gardens of 11
 Baghdad, as center of learning 83
 ballistics computations, and computer development 61–62
 base 2, counting in and computer design
 45, 55
 Leibniz and 45
 base 5, counting in examples of 9
 Malagasy method modified for 8
 base 10, counting in 8
See also decimal system in Babbage's analytical engine 51
 in Chinese method 30
 in early computers 54–55
 in Egyptian method 22
 examples of 9
 in Malagasy method 7
 superiority of 34
 base 20, counting in examples of 9
 in Mayan method 27–28, 28

base 60, counting in advantages of
 17–20
 in Mesopotamian method 14
 modern use of 40
 Bernoulli, Daniel 97
 Bernoulli, Jacob 97
 Bernoulli, Nicolaus II 97
 Bhaskara II (Hindu mathematician) 82
 on infinity 136–137
 binary notation
 in computers 45, 55
 Leibniz and 45
 Bolyai, János 104
 Bolzano, Bernhard 136–137
 study of infinite sets 138–140, 139, 141–142
 books
 Chinese, burning of 30, 32
 Mayan 26, 27
 Brahmagupta (Hindu mathematician)
 78–84
 concept of number 79
 concept of zero 80–81
 influence on Islamic mathematicians 83
 rules for subtraction 79–80
 Brahmagupta (Indian mathematician)
 Brāhma Sphuta Siddhānta 83
 Brāhma Sphuta Siddhānta (Brahmagupta) 78–79, 83
 concept of number described in 79
 Buddha 36
 Bush, Vannevar 52–53

C

calculator(s)
 mechanical:
 Leibniz's 50
 limitations of 48
 Pascal's 47–50
 Schickard's 47, 49
 programmable:
 algorithm used to compute square roots 72
 errors generated by 58–59
 calculus
 Leibniz and 45
 Wallis and 95
 Cambridge University, Turing at 174–175
Canon Mathematicus (Viète) 41
 Cantor, Georg 112, 143, 144
 and arithmetic of the infinite 151–153
 contributions of 163–164
 and Dedekind 144
 and degrees of infinity, speculations about 180
 and language of sets 154–155
 paradox discovered by 155–156
 theory of infinite sets 144–154
 Cardano, Girolamo 89
 Ars Magna 88, 91–92
 and Ferrari 90–91
 and Tartaglia 90, 91
 cardinal arithmetic 151–153, 153
 cardinality
 definition of 145
 of finite sets 151
 cardinal numbers, associated with infinite sets 150–151

cartography, use of complex numbers in 100

central processing unit (CPU), Babbage's terminology for 51

Chinese mathematics 30–32 fractions in 31–32 negative numbers in 77 rod numerals in 30–31, 31, 77

choice, axiom of 162–163

clay tablets, cuneiform writing on 12, 12–14

Codex Tro-Cortesianus 26

Collège des Quatre Nations 101

Colossus (computers) 60

common fractions, definition of 37

comparing *See also* correspondence vs. counting 3 sizes of finite sets 134, 140–141

complex numbers 70 d'Alembert on 101–102 applications of 100, 104–105 discovery of 85 Euler on 99, 100, 102 geometric representation of 102–106, 103 imaginary part of 103, 104 Leibniz on 96 real part of 102–103, 104 resistance to concept of 96–97 Wallis's geometric description of 96

complex number system description of, Dedekind's ideas and 115 discovery of 111

complex variables, graphs of functions of 106

composite number 124

computer(s) and approximation 56–57, 72 base 2 numeration (binary notation) in 45, 55 base 10 numeration (decimal notation) in 54–55 early design ideas for 52–53 first-generation 54–55, 55, 60–62 forerunner of (Babbage's analytical engine) 50–54, 51 and halting problem 178 limitations of 48, 56, 59 motivation for development of 54, 60–62 storing numbers in 57–63 Turing's contribution to development of 174, 175, 175–177 computer-programming languages 173 continuum hypothesis 153 generalized 154

correspondence and concept of number 5 of finite sets 140, 141 idea of 134–135

of infinite sets: Bolzano's observations about 138–140, 139, 141–142 Cantor's observations about 145–147 Galileo's observations about 134, 135, 141–142 tale of Tristram Shandy and 141

Leonardo's odometer and 4, 5

counting *See also under* base Chinese method of 30–31, 31 vs. comparing 3 in comparing sizes of finite sets 134, 140–141 cultures without method of 8–9 deceptive simplicity of xi development of methods of 6–7, 9–10 Egyptian method of 22 history of xi Malagasy method of 6–7 Mayan method of 8, 26–29, 28 Mesopotamian method of 8, 14–17, 36 modern system of, Indian mathematicians and 36–37 counting board 35 CPU (central processing unit), Babbage's terminology for 51 Cromwell, Oliver 95 cuneiform writing 11, 14

clay tablets with 12, 12–14
decoding of 12
number symbols in 16

D

d'Alembert, Jean Le Rond. *See* Alembert, Jean Le Rond d'
decimal comma 43
decimal fractions
introduction of 40–43
notations for 42
decimal point, introduction of 42, 43
decimal system
advantages of 46–47
development of 40–43
vs. floating-point notation 64

Dedekind, Richard 111–112
and Cantor 144
and definition of infinite sets 145
and real number line 112–113
on real number system 112–115

Dedekind cut 114
“The Definition of Number” (Russell) 6
degree of equation 86
and number of roots 94

del Ferro, Scipione 89–90
De Morgan, Augustus 97
De Numero Indorum (Latin translation of al-Khwārizmī) 83
Descartes, René
on negative numbers 85
on roots of algebraic equations 95

Dialogues Concerning Two New Sciences (Galileo) 131–132

Dichotomy, Zeno's 121–122

differential equations, use of complex numbers in 105

Diophantus of Alexandria 76

division

positive rational numbers as closed system under 69, 71

real numbers as closed system under 92
by zero 80

Brahmagupta on 80–81

Wallis on 81

Dresden Codex 26, 27

E

e (number) 108

Euler's study of 107–108

notation for 100

as transcendental number, proof of 110

education,

Mesopotamian 13

EDVAC (Electronic Discrete Variable Automatic Computer) 61

Egyptian mathematics 20–26

counting method in 22

fractions in 22–26

hieratic script in 22, 23

hieroglyphic notation in 21, 22

Einstein, Albert 171

Electronic Discrete Variable Automatic Computer (EDVAC) 61

Electronic Numerical Integrator and

Computer (ENIAC) 54–55, 55, 61

Elements (Euclid) 123, 124

axioms and postulates used in 160

proof of infinitely many prime numbers 124, 125–126

Tartaglia's translation of 89

title page of 123

ENIAC (Electronic Numerical Integrator and Computer) 54–55, 55, 61

Enigma machine (computer) 60

equation(s) *See also* algebraic equations and concept of number, development of 85, 87–88

degree of 86

Euclid of Alexandria 123–124

axiom on whole and part, Cantor's rejection of 144–145

axioms and postulates used by 160

Elements 123, 124

Tartaglia's translation of 89

title page of 123

Hilbert's contributions to geometry of 166

proof of infinitely many prime numbers 124, 125–126

Eudoxus of Cnidus 128

Euler, Leonhard

97–100, 100

on complex numbers 99, 100, 102

study of *e* number 107–108

Euler, Paul 97
 European mathematics
 concept of infinity in 129
 Hindu mathematics and 35, 37–40, 83
 resistance to negative numbers 85
 sexagesimal system used in 40
 even numbers, listing 125
 exhaustion, method of 128–129
 exit condition 128–129
 exponent 57–58
 extraction of roots 69
See also square root(s)
 numbers resulting from 70

F

Ferrari, Lodovico 90–91
 Fibonacci, Leonardo
See Leonardo of Pisa
 finite sets
 cardinality of 151
 comparing sizes of 134, 140–141
 Fior, Antonio Maria 90
 five *See* base 5
 fixed-point notation
 Babbage's system of 51–52
 limitations of 52–54, 56
 floating-point notation 56–64
 advantages of 63
 vs. decimal system 64
 fluid dynamics, use of complex numbers in 104–105
 Fontana, Niccolò *See* Tartaglia
 formal language(s)
 Russell and
 Whitehead's 168
 today 173

formal mathematics
 Hilbert's concept of 166–167
 impossibility of, Gödel on 171–172
 fractions *See also specific types of fractions*
 absence in Mayan mathematics 29
 in Chinese mathematics 31–32
 decimal:
 introduction of 40–43
 notations for 42
 in Egyptian mathematics 22–26
 in Hindu mathematics 37
 in Mesopotamian mathematics 17–20
 sexagesimal system used for 40
 function(s)
 of complex variables, graphs of 106
 in differential equations 105
 graphing of 106
 language of sets and 154
 functional analysis 165
 fundamental theorem of algebra 94

G

Galilei, Vincenzo 130
 Galileo Galilei 130, 131
 conclusions about infinite sets 135, 148
 contributions of 130–131
Dialogues Concerning Two New Sciences 131–132
 on perfect squares and natural numbers, relationship between 132–135, 133

Gauss, Carl Friedrich 104
 generalized continuum hypothesis 154
 geometric representation
 of complex numbers 102–106, 103
 of roots of quadratic equations 95–96
La Géométrie (Descartes) 95
 geometry
 Euclid's, logical flaws in 166
 and interpretation of numbers 67–68
 Lobachevsky's 161
 Girard, Albert 93–94
 Gödel, Kurt 170, 171
 incompleteness theorem of 171–173
 Göttingen University 111
 Department of Mathematics 112
 Greek mathematics
 axioms in 160, 161
 and concept of infinity 120–129
 concept of number in, vs. Hindu concept 84
 and irrational numbers 76
 numbering system in 33
 Pythagoreans and 73–76
 Grolier Codex 26

H

halting problem 178–179
 Hammurabi Code 11
 Han dynasty (China) 32
 Hanging Gardens of Babylon 11

heat conduction, and use of complex numbers 105
 Hermite, Charles 110
 hieratic numerals, Egyptian 22, 23
 hieroglyphic notation, Egyptian 21, 22
 Hilbert, David 165
 and axioms, questions of completeness and consistency of 167–169
 and Euclidean geometry, completely rigorous foundation for 166
 formalized mathematics of 166–167, 173
 impossibility of task of, Gödel on 171–172
 problems posed by 165–166
 Hilbert spaces 165
 Hindu mathematics 36–37
 adoption in Europe 35, 37–40, 83
 adoption in Islamic world 37, 83
 algebraic equations solved in 87
 characteristics of 82
 concept of infinity in 136–137
 history of 82
 negative numbers in 78–84, 81
 numeration system in, explanation of 38–39
 poetry in 36, 82

I

i (number), notation for 99

imaginary numbers 85
See also complex numbers
 first description of xii
 Wallis on 98–99
 imagination, human vs. machines 48
 self-imposed limitations on 70, xii
 incompleteness theorem, Gödel's 171–173
 Indian mathematics *See* Hindu mathematics
 infinite-dimensional spaces, theory of 165
 infinite processes
 Greek concept of 128–129
 vs. infinite sets 127
 infinite sets 130
 Bolzano's study of 138–140
 Cantor's definition of 145
 Cantor's theory of 144–154
 cardinal numbers associated with 150–151
 of different sizes, Cantor's demonstration of 148–149
 Galileo's observations about 134, 135, 141–142
 vs. infinite processes 127
 power sets of 150
 proper subsets of, equality of size of 134, 135, 145
 size of, transfinite numbers representing 151
 infinity. *See also* infinite processes; infinite sets
 absence of physical analog for 119
 Aristotle on 128
 Cantor's interest in 144
 degrees of, Cantor's speculations about 180
 difficulty in accepting 120
 Euclid's contribution to 124
 in European mathematics 129
 in Greek mathematics 120–129
 in Hindu mathematics 136–137
 importance of 120
 in Islamic mathematics 129
 questions about xii–xiii
 Zeno's paradoxes and 121–122
 Internet, invention of 60
Invention nouvelle en l'algèbre (Girard) 93
 irrational numbers 70–72
 as algebraic numbers 107
 decimal expansions of 112
 definition of 70, 111, 112
 discovery of 73, 74
 theoretical implications of 76
 first description of xii
 Mesopotamian approach to 70–72
 Pythagoreans on 74–76, 111
 and rational numbers, Dedekind on relationship of 114
 Islamic world
 algebraic equations solved in 87
 concept of infinity in 129

Hindu mathematics
adopted in 37, 83

Italy, Renaissance, discoveries in 87, 88–92

J

John Hollywood (John of Halifax). *See* Sacrobosco

K

Kepler, Johannes 47

The Keys to Mathematics (Oughtred) 95

al-Khwārizmī,
Mohammed ibn-Mūsā 37, 38, 83

King's College,
Cambridge University,
Turing at 174–175

Kronecker, Leopold 143

L

Lambert, Johann Heinrich 109

laws, first written code of 11

laws of nature, and differential equations 105

Legendre (Le Gendre), Adrien-Marie 109

Leibniz, Gottfried Wilhelm 43–45
and base 2 numeration 45
on complex numbers 96
mechanical calculator designed by 50

Leibniz wheel 50

Leonardo da Vinci
odometer designed by 3–5
self-portrait of 5

Leonardo of Pisa
(Leonardo Fibonacci) 37–39

Liber Abaci (Leonardo of Pisa) 38–39

The Life and Opinions of Tristram Shandy, Gentleman (Sterne) 140–141, 142

Liouville, Joseph 109–110

Lobachevsky, Nikolai Ivanovich 161, 161–162

logarithms, invention of 43

logic
mathematical,
Turing's perspective on 174–175

mathematics as branch of 169

M

MacLaurin, Colin 85

Madagascar, counting method used in 6–7

Malagasy method of counting 6–7
base 5 modification of 8

mantissa 57

Mas d'Azil (Ariège, France), pebbles excavated from 4

mathematics
as branch of logic 169
formal:
Gödel on impossibility of 171–172
Hilbert's concept of 166–167, 173
foundations of:
Russell's approach to 169–170
Turing's halting problem and 178–179

modern, beginning of 88

nature of, Gödel's incompleteness theorem and 172–173

religion and 45, 144

vs. science 160

Mauchly, John W. 61, 62

Mayans 26
counting method of 8, 26–29, 28

Dresden Codex 26, 27

mechanical calculator(s)
Leibniz's 50
limitations of 48
Pascal's 47–50
Schickard's 47, 49

memory
computer, Babbage's terminology for 51
in Turing machine 176

Mesopotamian civilization 11–12
algebraic equations solved in 87
counting method of 8, 14–17, 36
cuneiform writing of 11, 12, 12–14
education in 13
fractions used in 17–20
and irrational numbers, approach to 70–72
number symbols in 16
“problem text” from 18

metric system 42

Mirifici Logarithmorum Canonis Constructio (Napier) 43, 44

modern mathematics, beginning of 88

modern number system,
development of 34–45
Moors 37
morphogenesis, Turing's
contribution to 174
Moscow papyrus 21
Mozart, Wolfgang
Amadeus 97
multiplication
Egyptians and 25
positive rational num-
bers as closed system
under 69, 71
real numbers as closed
system under 92
transfinite 152–153,
153

N

Napier, John 40, 42–43
Mirifici Logarithmorum Canonis Constructio
43, 44
notation for decimal
fractions 42
Napier's rods (bones) 43
natural numbers
and algebraic num-
bers, correspondence
between 146–147
inability to count 134
and perfect squares,
Galileo's observa-
tions about
132–135, 133
power set of, Cantor
on 150
and rational numbers,
Cantor's observations
about 145–147
and real numbers,
Cantor's proof
regarding 148–149
negative numbers
77–84
in Chinese mathemat-
ics 77

development of
understanding of 67
European mathemati-
cians and 85, 96–97
Girard's understand-
ing of 93
Indian mathemati-
cians on 78–84, 81
indications about exis-
tence of 69–70
representation on real
number line 68, 77
resistance to concept
of 69, 70
square roots of 92,
104
Wallis on 98–99
Neumann, John von 55
Newton, Isaac, Wallis's
influence on 95
*The Nine Chapters on the
Mathematical Art*
(Chinese text) 31–32
problem from 32
non-Euclidean geom-
etry
Bolyai and 104
Lobachevsky and 161
nonexistence proof 75
Nova Scientia (Tartaglia)
89
number(s) *See also spe-
cific kinds of numbers*
absolute value of 105
Brahmagupta's con-
cept of 79
classification scheme
for 111
concept vs. example
of 6
current inquiries into
180
early concepts of 67,
68, 69
evolution of concept
of 9, 33, 179
algebraic equations
and 85, 87, 91, 93
Cantor and 151
Girard's concept of
93
Hindu system of,
explanation of
38–39
Hindu vs. Greek con-
cept of 84
al-Khwārizmī's con-
cept of 83
modern system of,
development of
34–45
new kinds of, difficul-
ty in accepting xi–xii
19th-century under-
standing of, limita-
tions of 112–113
prehistoric ideas
about 3, 4
representation on real
number line 67–68
Russell's definition of
6
number system
Cantor's extension of
151
modern, development
of 34–45
numerical algebra 20

O

odometer, early design
for 3–5
one (number) 124
Ordinance Variable
Automatic Computer
(ORDVAC) 62
Oughtred, William 95

P

papyrus, Egyptian writ-
ing on 20–21
Paramenides 120, 121,
122
Pascal, Blaise
influence on comput-
er design 53

mechanical calculator
designed by 47–50

Pascalines 47–50

Perez Codex 26

perfect square(s)
definition of 132
and natural numbers,
Galileo's observations about
132–135, 133

philosophy,
Pythagorean 73–74

π (number) 108
study of 109–110
as transcendental number, proof of 110

Pilot ACE computer 175

Plato, irrational numbers and 75

poetry, Hindu mathematics and 36, 82

positional numeration
absence in Egyptian hieratic system 22
Hindu system of 37
Mayan system of 28, 29
Mesopotamian system of 15, 17
superiority of 34

positive numbers, transition to algebraic numbers from 92

positive rational numbers 67
closed system of, conditions for 69, 71
numbers different from 70

power sets 150

prime numbers
definition of 124
inability to list all 124–125
infinitely many, Euclid's proof of 124, 125–126

Principia Mathematica (Russell and Whitehead) 156–157, 168, 169–170
and computer programs, similarities between 173

Gödel's work based on 172

programmable calculators
algorithm used to compute square roots 72
errors generated by 58–59

Project PX 61

proof
absence in Hindu mathematics 82
by contradiction 126–127
nonexistence 75

proper subset
definition of 132
of infinite set, equality of size of 134, 135, 145

Pythagoras of Samos 73, 73–74

Pythagoreans 73–76
on irrational numbers 74–76, 111
philosophy of 73–74

Pythagorean theorem
Chinese use of 32
Mesopotamian version of 18

Q

Qin Shi Huang (emperor of China) 30
book burnings ordered by 30, 32

quadratic equations 86, 87

geometric representation of roots of 95–96

quadratic formula 90

R

rational number(s)
as algebraic numbers 107
decimal expansions of 112
definition of xii
denseness on real number line 71–72
and irrational numbers, Dedekind on relationship of 114
and natural numbers, Cantor on relationship of 145–147

real number(s)
closed system of, conditions for 92
definition of xii
infinite sets of, Bolzano's ideas about 138, 138–140, 139
as larger set than natural numbers, Cantor's proof of 148–149

real number line 67–68
Dedekind's contribution to 112–113
Girard's anticipation of 93
negative numbers on 68
rational numbers on 71–72
Wallis's conception of 95–96

real number system, Dedekind on 112–115

recursive algorithm 71

religion, and mathematics 45, 144

Renaissance, algorithms discovered during 86, 87, 88–92
 Rhind, Alexander Henry 21
 Rhind papyrus *See* Ahmes papyrus
 rod numerals, Chinese 30–31, 31, 77
 roots, extraction of 69
See also square root(s)
 numbers resulting from 70
 roots of algebraic equations 87, 107
 Descartes on 95
 Girard on 94
 Wallis on 95–96
 round-off error
 computers and 57
 Stevin on 41
 Russell, Bertrand 155, 156, 165
 approach to foundations of mathematics 169–170
 “The Definition of Number” 6
 Principia Mathematica 156–157, 168, 169–170
 and tale of Tristram Shandy 140
 Russell paradox 157–159
 as logical error 169
 solution to 162

S

Sacrobosco, Johannes de 38–39
 Saint Petersburg Academy, Russia 97
 Schickard, Wilhelm, calculator designed by 47, 49
 science, vs. mathematics 160

Sebokt, Severus 37
 set(s)
 concept of, introduction of 140
 finite:
 cardinality of 151
 comparing sizes of 134, 140–141
 infinite *See* infinite sets
 language of, universality of 154–155
 problem of defining 157–158
 universal, disproving existence of 159
 well-ordering of 165–166
 set theory, Cantor’s 144–154
 functions studied in context of 165
 importance of 163–164
 paradox in 155–156, 157–159
 Russell and Whitehead’s reinterpretation of 170
 Russell’s studies in 157
 Zermelo’s contribution to 162–163
 Seven Wonders of the Ancient World 11
 sexagesimal system 14
 advantages of 17–20
 fractions represented with 40
 in Mesopotamian mathematics 14
 modern use of 40
 Shih Huang-ti (emperor of China) *See* Qin Shi Huang (emperor of China)
 Sidhanta (Hindu astronomical text) 37
 sine function, origins of 36
 sixty *See* base 60
 square root(s)
 algorithm used by calculators to compute 72
 of negative numbers 92, 104
 square root of 2
 algebraic equation associated with 110
 description of, Dedekind’s ideas and 115
 irrationality of:
 geometric representation of 74
 proof of 75–76
 Mesopotamian computation of 20, 71, 72
 number resulting from 70
 Pythagoreans on 73
 square root of -2, number resulting from 70
 Sterne, Laurence, *The Life and Opinions of Tristram Shandy, Gentleman* 140–141, 142
 Stevin, Simon 40, 41–42
 notation for decimal fractions 42, 42
 subtraction
 Brahmagupta’s rules for 79–80
 and negative numbers 69–70
 real numbers as closed system under 92

T

Tartaglia (Niccolò Fontana) 88, 88–89
 algorithm for third-degree algebraic equations 89–90
 and Cardano 90, 91

ten *See* base 10; decimal system
The Tenth (Stevin) 41–42
 theorems 160
 transcendental numbers best known 108 criteria for 108–109 definition of 107 difficulty in describing 110 Euler on 107–108 number of, conclusion regarding 150 proof of existence of 109–110 transfinite numbers arithmetic of, Cantor and 151–153, 153 definition of 151 questions about 153 transistors, and computer design 54
Trattato di Numeri et Misure (Tartaglia) 89
Treatise of Fluxions (Maclaurin) 85
A Treatise on Algebra (Wallis) 98–99
 Tübingen University, reconstruction of Schickard's calculator at 47, 49
 Turing, Alan 174–175
 Turing machine 175–177 computer modeled on 175 diagram of 176 and halting problem 178
 twenty *See* base 20
 two *See* base 2

U

Ujjain (India), astronomical observatory at 78, 78, 136

United States Army, and computer development 61–62
 United States Census Bureau, computer created for 62 unit fractions, Egyptians and 24–25
 Universal Automatic Computer (UNIVAC) 62 universal set, disproving existence of 159 universe, finite nature of 119, 119–120
 University of Basel, Euler at 97
 University of Berlin Cantor at 143 Zermelo at 162, 163
 University of Freiburg, Zermelo at 163
 University of Göttingen, Hilbert at 165
 University of Halle, Cantor at 143
 University of Illinois, computer development at 62
 University of Königsberg, Hilbert at 165
 University of Pennsylvania, computer development at 61
 University of Piza, Galileo at 130
 University of Prague, Bolzano at 136, 137
 University of Vienna, Gödel at 170

V

vacuum tubes, and computer design 54
 Viète, François 40–41 notations for decimal fractions 41, 42

vigesimal system 28
 Mayan 27–28, 28

W

Wallis, John 94, 95–96 on division by zero 81 influence of 95 and quadratic equations, geometric representation of roots of 95–96
A Treatise on Algebra 98–99
 war effort, and computer development 54, 60–62
 Weierstrass, Karl 143 well-ordering of sets 165–166
 Wessel, Caspar 103
 Whitehead, Alfred North, *Principia Mathematica* 156–157, 168, 169–170
 World War II and computer development 54, 60–61 Turing's role during 174
 writing Mesopotamian method of 11, 12 and numbers, use of 3

Z

Zeno of Elea 120 paradoxes of 120–122
 Zermelo, Ernest 162–163, 166
 zero Brahmagupta on 80–81

Chinese symbol for 31
discovery of symbol
for 17, 36
division by 80
Brahmagupta on
80–81
Wallis on 81

Hindu symbol for 37
indications about exis-
tence of 69
introduction of,
impact of 34
Mayan symbol for
28, 29

Mesopotamian
version of 17
need for symbol for
15–16
resistance to concept
of 69, 70

